

Using Integrated Administrative Data to Identify Youth Who Are at Risk of Poor Outcomes as Adults

Prepared by Keith McLeod, Robert Templeton, Christopher Ball
Sarah Tumen, Sarah Crichton and Sylvia Dixon for the
Treasury

Analytical Paper 15/02

December 2015



DISCLAIMER

The views, opinions, findings, and conclusions or recommendations expressed in this Analytical Paper are strictly those of the author(s). They do not necessarily reflect the views of the New Zealand Treasury, Statistics New Zealand or the New Zealand Government. The New Zealand Treasury and the New Zealand Government take no responsibility for any errors or omissions in, or for the correctness of, the information contained in this Analytical Paper.

The results in this report are not official statistics, they have been created for research purposes from the Integrated Data Infrastructure (IDI) managed by Statistics New Zealand.

Access to the anonymised data used in this study was provided by Statistics NZ in accordance with security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business or organisation and the results in this paper have been confidentialised to protect these groups from identification.

Careful consideration has been given to the privacy, security and confidentiality issues associated with using administrative and survey data in the IDI. Further detail can be found in the Privacy impact assessment for the Integrated Data Infrastructure available from www.stats.govt.nz.

The results are based in part on tax data supplied by Inland Revenue to Statistics NZ under the Tax Administration Act 1994. This tax data must be used only for statistical purposes, and no individual information may be published or disclosed in any other form, or provided to Inland Revenue for administrative or regulatory purposes.

Any person who has had access to the unit-record data has certified that they have been shown, have read, and have understood section 81 of the Tax Administration Act 1994, which relates to secrecy. Any discussion of data limitations or weaknesses is in the context of using the IDI for statistical purposes, and is not related to the data's ability to support Inland Revenue's core operational requirements.

ANALYTICAL PAPER 15/02	Using Integrated Administrative Data to Identify Youth Who are at Risk of Poor Outcomes as Adults
MONTH/YEAR	December 2015
AUTHORS	Keith McLeod, The Treasury Robert Templeton, The Treasury Christopher Ball, The Treasury Sarah Tumen, The Treasury Sarah Crichton, The Treasury Sylvia Dixon, The Treasury
ISBN (ONLINE)	978-0-908337-33-0
URL	Treasury website at December 2015: http://www.treasury.govt.nz/publications/research-policy/ap/2015/15-02 Persistent URL: http://purl.oclc.org/nzt/p-1810
NZ TREASURY	New Zealand Treasury PO Box 3724 Wellington 6008 NEW ZEALAND Email information@treasury.govt.nz Telephone 64-4-472 2733 Website www.treasury.govt.nz
PURPOSE OF THE ANALYTICAL PAPERS SERIES	The Treasury's aim in publishing the Analytical Papers series is to make this analysis available to a wider audience and to inform and encourage public debate, with the ultimate aim of informing our policy advice. Analytical Papers are commissioned as part of the Treasury's core function in developing and providing advice to Ministers. They include work undertaken by staff at the Treasury or other government departments, as well as work undertaken for the Treasury by external researchers or consultants. Analytical Papers do not themselves represent policy advice.

© Crown Copyright



This work is licensed under the Creative Commons Attribution 4.0 International licence. In essence, you are free to copy, distribute and adapt the work, as long as you attribute the work to the Crown and abide by the other licence terms.

To view a copy of this licence, visit <https://creativecommons.org/licenses/by/4.0/>. Please note that no departmental or governmental emblem, logo or Coat of Arms may be used in any way that infringes any provision of the [Flags, Emblems, and Names Protection Act 1981](#). Attribution to the Crown should be in written form and not by reproduction of any such emblem, logo or Coat of Arms.

Contents

- Abstract 1
- Executive summary 2
 - Purpose of the paper 2
 - Research objectives..... 2
 - Data and methods..... 3
 - Limitations and caveats 3
 - Key findings 4
- 1 Introduction 5
 - 1.1 Purpose..... 5
 - 1.2 Outcome measures..... 5
 - 1.3 Approach to defining the ‘at-risk’ populations 6
 - 1.4 Report structure 6
- 2 Data and methods..... 7
 - 2.1 Data description and limitations 7
 - Service cost data..... 8
 - 2.2 Populations of interest..... 9
 - 2.3 Estimating future outcomes and costs using statistical matching 11
 - 2.4 Approach to the identification of target populations 12
 - Identifying risk factors and predicting risk 12
 - Defining and describing target populations 12
 - 2.5 General caveats and cautions..... 13
- 3 Which young people experience poor outcomes?..... 14
 - Socio-demographic characteristics 14
 - Location in New Zealand..... 15
 - Located overseas..... 16
 - Childhood risk factors..... 17
 - Schooling 19
 - Tertiary education 21
 - Work and welfare 22
 - Early corrections contact..... 24
 - Early use of mental health services 25
 - Early parenting 26

4	Predicting poor outcomes	28
4.1	Regression model factor selection and estimation.....	28
	Predictive factors selected in the models.....	28
	Model discrimination	30
4.2	Predicting risk and defining populations with high risk.....	31
	Risk across multiple poor outcomes – characteristics and expected outcomes	32
	Extreme risk of one or more outcomes – characteristics and expected outcomes.....	35
	Comparing risk measures	37
	How does estimated risk change over time?	38
5	Target populations	40
5.1	Approach.....	40
5.2	Target population descriptions and criteria	40
5.3	Target population overlap and coverage.....	41
5.4	Target population projected outcomes and costs	45
6	Interpreting the A3 document.....	47
6.1	Identifying poor long-term outcomes (page 1)	47
6.2	The risk factors most associated with those outcomes (15-year-olds) (page 1)	47
6.3	Identifying those most at risk (page 1)	48
6.4	Characteristics of at-risk groups by age (page 2)	49
6.5	Target populations 15 to 19 years and 20 to 24 years (pages 3 to 6).....	49
7	Conclusions.....	50
	Appendices.....	51
	Appendix 1: Predictive factors at different ages.....	51
	Appendix 2: Outcomes by territorial authority	53
	Appendix 3: Estimated risk by territorial authority, 2013 youth population.....	55
	Appendix 4: Distribution of projected costs by estimated risk of poor outcomes at age 20	57
	Appendix 5: Target populations by territorial authority, 2013 youth population	59
	Appendix 6: Target population overlaps – 1990/1991 birth cohort.....	65

Figures and tables

Table 1: Socio-demographic characteristics and outcomes for youth 1990/91 cohort.....	15
Table 2: Outcomes for youth 1990/91 cohort by whether located overseas at each year of age.....	16
Table 3: Childhood risk factors and outcomes for youth 1990/91 cohort at age 15	18
Table 4: Schooling characteristics and outcomes for youth 1990/91 cohort at age 18	20
Table 5: Tertiary education and outcomes for youth 1990/91 cohort at age 22	21
Table 6: Employment and welfare characteristics and outcomes for youth 1990/91 cohort at age 22.....	23
Table 7: Corrections contact and outcomes for youth 1990/91 cohort at age 22.....	25
Table 8: Use of mental health services for youth 1990/91 cohort at age 20	26
Table 9: Early parenting and offspring childhood risk factors for youth 1990/91 cohort at age 22.....	27
Table 10: Areas under receiver operating characteristic (ROC) curves for each youth outcome model	31
Table 11: Demographic characteristics of the December 2013 youth population by level of risk across multiple outcomes.....	32
Figure 1: Percentage of the December 2013 youth population at high risk across multiple poor outcomes by territorial authority area (Auckland territorial authorities expanded).....	33
Table 12: Estimated outcomes for the 1990/91 birth cohort population by level of risk across multiple outcomes.....	34
Table 13: Projected welfare and corrections costs aged 25 to 34 by estimated risk across multiple outcomes at age 20, 1990/91 birth cohort.....	34
Table 14: Demographic characteristics of the December 2013 youth population by whether at extreme risk of one or more outcomes	35
Table 15: Estimated outcomes for the 1990/91 birth cohort population by whether at extreme risk of one or more outcomes	35
Figure 2: Percentage of the December 2013 youth population at extreme risk of one or more poor outcomes by territorial authority area (Auckland territorial authorities expanded).....	36
Table 16: Projected welfare and corrections costs for ages 25 to 34 by whether at extreme risk of one or more outcomes at age 20, 1990/91 birth cohort	37
Table 17: Estimated multiple outcome risk and extreme risk of one or more poor outcomes at age 15, 1990/91 birth cohort.....	38
Table 18: Estimated multiple outcome risk and extreme risk of one or more poor outcomes at age 21, 1990/91 birth cohort.....	38
Table 19: Estimated multiple outcome risk and extreme risk of one or more poor outcomes at age 15, 1990/91 birth cohort.....	39

Table 20: Estimated multiple outcome risk and extreme risk of one or more outcomes at age 15, 1990/91 birth cohort.....	39
Table 21: Target population descriptions, criteria, size and estimated risk.....	42
Figure 3: Target population overlaps December 2013 population, ages 15 to 19.....	43
Figure 4: Target population overlaps December 2013 population, ages 20 to 24.....	44
Table 22: Expected outcomes by target population	45
Table 23: Projected costs by target population group – total from age 25 to 34	46
Appendix 1 Table 1: Predictive factors and number of models by age	51
Appendix 2 Table 1: Estimated outcomes at ages 25 to 34 for youth 1990/91 cohort by the territorial authority of residence at age 15	53
Appendix 3 Table 1: Estimated risk for December 2013 population aged 15 to 24 by territorial authority of residence	55
Appendix 4 Figure 1: Projected welfare and corrections cost distribution by risk level at age 20, 1990/91 birth cohort (numbers of people – \$0 to \$100,000 only)	57
Appendix 4 Figure 2: Projected welfare and corrections cost distribution by risk level at age 20, 1990/91 birth cohort (numbers of people – over \$100,000 only)	57
Appendix 4 Figure 3: Projected welfare and corrections cost distribution by risk level at age 20, 1990/91 birth cohort (percentage of each risk level population – \$0 to \$100,000 only)	58
Appendix 4 Figure 4: Projected welfare and corrections cost distribution by risk level at age 20, 1990/91 birth cohort (percentage of each risk level population – over \$100,000 only)	58
Appendix 5 Table 1: Target populations by territorial authority for December 2013 population (ages 15 to 19).....	59
Appendix 5 Table 2: Target populations by territorial authority for December 2013 population (ages 20 to 24).....	62
Appendix 6 Figure 1: Target population overlaps 1990/91 cohort, ages 15 to 19.....	65
Appendix 6 Figure 2: Target population overlaps 1990/91 cohort, ages 20 to 24.....	66

Abstract

This paper summarises findings from an analysis of integrated administrative data seeking to identify the characteristics of young people aged 15 to 24 who are most at risk of poor long-term outcomes. The research is part of a broader ‘social investment approach’ by government agencies seeking to target services more effectively towards those most at need and reflects the recognition that such an approach requires better evidence about who these at-risk groups are. The analysis identifies those characteristics in the administrative data that are most predictive of a range of future poor outcomes and how this changes over the course of a young person’s entry into adulthood and identifies groups of young people at particular risk at different ages.

JEL CLASSIFICATION

I38 Welfare, well-being and poverty:
Government programmes

J13 Childcare, children and youth

C55 Large data sets: modelling and analysis

Executive summary

Purpose of the paper

This paper summarises findings from an analysis of integrated administrative data aiming to identify the characteristics of young people aged 15 to 24 who are most at risk of poor long-term outcomes. The work is part of a broader emphasis by government agencies to target services more effectively towards those most at need and reflects the recognition that such an approach requires better evidence about who these at-risk groups are.

The Treasury has identified the need for the state sector to play a particular role in helping the most disadvantaged to participate in society and the economy and has noted the importance of agencies doing this through working innovatively and collaboratively across agency boundaries.¹ This has driven the development of a ‘social investment’ approach to decision making about government investment in social services. The social investment approach involves “using information and technology to better understand the people who need public services and what works, and then adjusting services accordingly”.²

This work represents one in a number of steps towards implementing a social investment approach. The overall work programme was led by the Ministry of Education, with the analysis summarised in this paper being led by the Treasury’s Analytics and Insights team, in collaboration with a number of other agencies, and using integrated administrative data held in Statistics New Zealand’s Integrated Data Infrastructure (IDI).

The results of this analysis are also described in the accompanying A3 document entitled ‘Youth at risk: Identifying a target population’, produced by the Ministry of Education. This paper provides a general description of the process adopted, presents a descriptive analysis of the populations of interest, summarises the results of the modelling work undertaken and describes the target populations identified through the project. It also provides some guidance to assist with interpreting the A3 document.

Research objectives

The aim of this work is to identify which risk factors between the ages of 15 and 24 are most strongly associated with poor long-term outcomes at ages 25 to 34, identify target populations between the ages of 15 and 24 who are most at risk of experiencing poor long-term outcomes and identify some of the larger fiscal costs associated with those target populations.

¹ <http://www.treasury.govt.nz/publications/briefings/holding-on-letting-go/>.

² <http://www.treasury.govt.nz/statesector/socialinvestment>

Data and methods

The study uses the Integrated Data Infrastructure (IDI), which brings together information from a wide range of government departments. Records are linked using name and date of birth. The data is anonymised and used only for research purposes.

The main analysis is a birth cohort analysis, which focuses on those born between 1 July 1990 and 30 June 1991, who can be observed through to age 21 in the data set. The analysis describes the key characteristics and outcomes that could be observed for this cohort at various ages and the various service use patterns and outcomes that were experienced by different subgroups within this population. The future outcomes of this birth cohort out to age 35 are also estimated using a statistical record linkage technique, in which data for an older birth cohort is linked to that of the 1990/91 cohort.

This is complemented by an analysis of the characteristics and outcomes of the current youth population, defined as being aged 15 to 24 as at the end of December 2013. We are able to describe these young people's interactions with selected social services up to the end of December 2013. Projected future outcomes and selected service costs are also estimated for this population using data for other birth cohorts and statistical record linkage techniques.

Limitations and caveats

The study has a number of limitations and caveats:

- ▶ The scope of the study is limited by the nature and breadth of the information collected in agencies' administrative systems and included in the IDI. For example, the administrative data used in this work provides only a partial picture of childhood adversity, service use and service costs.
- ▶ The population coverage errors, linkage errors and biases present mean that the results are unlikely to be completely accurate and should be viewed as providing broad estimates of scale.
- ▶ The methods used to estimate future outcomes and costs are designed to provide a comparative picture of future outcomes and costs for different population subgroups, but they have some significant limitations. These estimates should *not* be viewed as forecasts of the actual outcomes and costs that will be incurred in the future.

While the results highlight the power of using integrated administrative data in new and innovative ways, this is the first time some of the data has been used in this way, and as such, these results should be considered as preliminary and will need further testing and development over time.

The caveats and limitations are discussed in more detail later in the paper.

Key findings

- ▶ Integrated administrative data can be a powerful tool for government and other agencies to identify at-risk groups in the population. Limitations in some of the data mean that the findings of this analysis need to be treated with some caution. However, the results provide a useful insight into the lives of at-risk youth. The data used for this type of analysis will continue to improve over time.
- ▶ A number of characteristics can be identified throughout a person's early life that are predictive of future poor outcomes including early contact with government agencies such as Child, Youth and Family (CYF), demographic characteristics and geographic location, characteristics of the young person's caregiver and early outcomes evident in data from the education, corrections, welfare and health systems. These can be used to quantify risk at an individual level and to identify the size and characteristics of at-risk groups of young people at different ages.
- ▶ The characteristics that are predictive of future outcomes change over time. As young people progress into early adulthood, poor future outcomes become directly evident through contact with the benefit, corrections and health systems. Whilst it becomes easier to predict poor outcomes as a young person ages, these outcomes may become more difficult to influence.
- ▶ It is possible to identify groups of at-risk youth at different ages using a small set of identifying characteristics. However, these predictions are by no means perfect. Young people who are identified as being at risk are highly likely to have poor future outcomes, but a large number of people have poor outcomes despite not falling into one of these defined groupings.
- ▶ In general, geographic location is strongly associated with risk of poor outcomes, with location-based measures such as the New Zealand Deprivation Index (NZDep) and territorial authority area being important predictors of risk, even controlling for other observed characteristics. Youth at risk of poor outcomes tend to be concentrated in specific areas such as the Far North, Kawerau, Opotiki and Wairoa. However, it is important to note that the largest numbers of at-risk youth still live in larger urban centres such as Manukau, Waitakere, Hamilton and Christchurch.

1 Introduction

1.1 Purpose

This work makes use of integrated administrative data collected from across a number of government agencies in the Integrated Data Infrastructure (IDI) and managed by Statistics New Zealand to ensure the security and confidentiality of people's information. The analysis is part of a broader direction by government agencies to work in a more collaborative way to better target social services where they are most needed. This study is focused on the youth population, defined as people aged 15 to 24, and seeks to identify specific groups within this population that are at particular risk of poor longer-term outcomes. The outcome measures used in the study and the approach to measuring risk are discussed below.

1.2 Outcome measures

A set of measures were identified from data available in the IDI, and based on the Youth Outcomes Framework,³ across the domains of:

- ▶ Enjoying Economic Opportunity
- ▶ Engaging and Achieving in Education
- ▶ Maintaining Good Health
- ▶ Enjoying Safety and Security.

Outcomes considered in the analysis were derived from educational, health, corrections and welfare outcomes data. The measures used were considered to be the best that could be developed in the time available and represent a broad, but not exhaustive, range of poor outcomes that negatively impact on the lives of young people. In many cases, improved measures are likely to be able to be developed as the data in the IDI continues to develop and expand. Measures were defined as follows:

- ▶ Not achieving:
 - at least an NCEA level 2 qualification by age 23 (for those turning 15 or 16)
 - at least a level 4 qualification by age 23 (for those turning 17 to 21).
- ▶ Use of mental health or addiction services between ages 20 and 22 inclusive (for those turning 15 to 20).
- ▶ Receiving a custodial or community sentence between ages 25 and 34 inclusive.
- ▶ Being on a benefit for five years or more between ages 25 and 34 inclusive.

³ This framework includes five domain areas (the four listed above as well as Social Participation, for which an appropriate measure proved difficult to identify in the IDI). It was derived from the Global Youth Wellbeing Index, which set out six domains by which youth wellbeing could be defined – equivalents to the five used in the Youth Outcomes Framework as well as Information and Communication Technology. Information about the Global Youth Wellbeing Index can be accessed at www.youthindex.org.

The preferred approach was to analyse youth outcomes at ages 25 to 34 when possible. The 'mental health or addiction services' outcome measure was measured at ages 20 to 22 simply because the data needed to estimate it at ages 25 to 34 was not available. Over time, historical data will become more complete in the IDI, and such compromises will not be necessary.

1.3 Approach to defining the 'at-risk' populations

In order to define target 'at-risk' populations, a multi-stage process was adopted that:

1. identified the factors most associated with the outcomes of interest, calculated estimated risk scores for each individual and identified an 'at-risk' population
2. identified groups of people (or 'clusters') with similar identifying characteristics within this 'at-risk' population
3. defined and described a small set of proposed target populations that broadly matched the identified clusters.

The at-risk groups and associated target populations are then able to be described according to their level and type of risk, key predictors of risk and associated fiscal welfare and corrections costs.

Unless otherwise stated, people were identified on their birthday when they turned 15 to 22, so characteristics for 15-year-olds, for example, covered the time leading up to that birthday when they were 14 or younger. Characteristics at age 22 were used for ages 23 and 24.

A person was considered to be 'at risk' where they were in the top 5% estimated risk group for at least one of the four outcomes identified above. An additional multiple risk measure was also used based on a person having high risk across multiple outcomes measures. This was defined by ranking the estimated risk scores of each of the four outcomes, averaging these ranks for each person and then taking the 15% of the population considered to be most at risk according to their average rank.

1.4 Report structure

The structure for the rest of the paper is outlined below:

- ▶ Section 2 describes the data and methods used, including any limitations and caveats.
- ▶ Section 3 describes the expected outcomes for young people with different characteristics.
- ▶ Section 4 describes the results of the regression modelling exercise undertaken to identify the key predictors of poor outcomes for the youth population and describes the at-risk populations.
- ▶ Section 5 describes the characteristics and definition of the target populations identified by the project.
- ▶ Section 6 provides some guidance in interpreting the accompanying A3 document.
- ▶ Section 7 concludes.

2 Data and methods

2.1 Data description and limitations

The study uses the Integrated Data Infrastructure (IDI), which was developed and is maintained and held by Statistics New Zealand. The data is held in a secure environment and made available to bona fide researchers under strict conditions. The IDI includes a wide range of survey and administrative data from across government agencies. This study uses data sourced primarily from the Ministry of Social Development (MSD – related to benefits, CYF care and protection and youth justice), the Department of Corrections (sentencing), the Ministry of Education (schooling and tertiary study participation and achievement), the Department of Internal Affairs (birth and death registrations), the Ministry of Health (mental health and addiction service usage and mental health pharmaceuticals), Inland Revenue (salaries and wages) and the Ministry of Business, Innovation and Employment (movements into and out of New Zealand).

Data is rounded to a multiple of three to protect confidentiality, and small cells are suppressed. As a result data in tables and figures may not add exactly to totals.

A number of potential data issues were outlined in Treasury Analytical Paper 15/01,⁴ and are relevant to this study also. Some of these issues are summarised here:

- ▶ The IDI includes information on children who were referred to CYF by the New Zealand Police because they had broken the law. However, the proportion of young people who have contact with the New Zealand youth justice system as a whole is higher than reported here. This is because the vast majority of apprehensions by the Police are dealt with by caution or warnings or by the Police Youth Aid Section and reflects the system's emphasis on diverting young offenders who commit lower-level offences away from formal youth justice processes where possible.
- ▶ Benefit data can be used to identify periods when children and young people are supported by a benefit as a child, periods when they are the primary recipient of a benefit in adulthood and periods when they are supported by a benefit as the partner or spouse of a primary benefit recipient. The benefit data covers the period 1 January 1993 onwards, so for the 1990/91 cohort, this only observes whether the child is supported by benefits after around age two.
- ▶ Benefit data has been used to identify the parents and caregivers of the young people in the study populations, which means that some information about parents or caregivers, such as their corrections sentencing history or their qualifications is only known for children who have been supported by a benefit at some stage. For the 1990/91 cohort, this is about 50% of young people.

⁴ Crichton, S., Templeton, R., and Tumen, S. (2015). *Analytical Paper 15/01: Using Integrated Administrative Data to Understand Children at Risk of Poor Outcomes as Young Adults*. The Treasury. See: <http://www.treasury.govt.nz/releases/2015-09-14>.

- ▶ A child or young person's contact with CYF for care and protection reasons can be divided into a number of different levels of contact⁵ depending on the *highest* level of intervention. Administratively derived measures of the proportion of children who have had a finding of abuse or neglect may not provide a reliable measure of the real occurrence of child maltreatment however. They will reflect both variations in the extent to which children who experience maltreatment are notified to CYF as well as the uncertainty inherent in making a determination that maltreatment has occurred. In addition, the CYF data is incomplete in the early 1990s, and therefore some of the estimates of service use for this period will be understated.

Service cost data

This paper includes estimates of some of the largest fiscal costs that are associated with different groups of individuals when they are aged 25 to 34. The costs included cover benefit payments and the costs of serving sentences administered by the Department of Corrections. Only these costs were analysed because our method of estimating costs at ages 25 to 34 uses data for an earlier birth cohort (those born in 1978/79), and information on the use of other government services is not available for that earlier cohort.

All cost estimates used in this study are CPI adjusted to December 2014 dollars.

Benefit costs

Benefit costs were categorised into three groups – Tier 1 (main benefits), Tier 2 (supplements) and Tier 3 (additional support for people in hardship). Working for Family tax credits, student allowances and student loans were not included in the study.

In the early childhood period of the study, the Tier 1 benefits included Domestic Purposes Benefit, Widow's Benefit, Unemployment Benefit, Sickness Benefit, Invalid's Benefit, Orphan's and Unsupported Child Benefits, Independent Youth Benefit and Emergency Benefit. More recently, following changes to the benefit system in 2013 (and some earlier changes), Tier 1 benefits have included Jobseeker Support, Sole Parent Support, Supported Living Payment and their subcategories as well as Youth Payment and Young Parent payment.⁶

Tier 2 supplements include Accommodation Supplement, Family Tax Credit (not including payments made by IRD), Disability Allowance, Orphan's Benefit or Unsupported Benefit, and Foster Care Allowance. Tier 3 additional supports included Funeral Grant, Special Needs Grant and Temporary Additional Support.

⁵ These categories are as follows:

- 'Notification' occurs where a member of the public or an agency has expressed a concern about the care or protection of the child to CYF and this has been recorded as a report of concern by a social worker. This includes cases where no abuse or neglect is substantiated.
- 'Substantiated findings of abuse or neglect' occur where a social worker has made a formal finding that the child has suffered abuse or neglect. This may also include cases where there is a Family Whānau Agreement or Family Group Conference but no care episode recorded.
- 'Care' occurs when a court has determined that a child or young person is in need of care and protection and grants a custody or guardianship order. In most cases, the child or young person will have had a substantiated finding of abuse or neglect.

⁶ Note that Emergency Benefit, Orphan's Benefit and Unsupported Child's Benefit were unchanged.

Corrections costs

The costs of serving custodial and community sentences were calculated by multiplying the length of each sentence (taking the days actually served) by an average cost per day from a table of average per day sentence costs provided by the Department of Corrections.⁷ The sentences and orders for which cost data is available include prison, remanded in custody, supervision and related sentences (including extended supervision orders and intensive supervision), community detention, community work, other community sentences, home detention, parole, post-detention conditions, released to home detention and released with conditions.

The average cost figures provided by the Department of Corrections related to the last four financial years. In this analysis, the cost figures for those four years were averaged (giving more weight to recent data) and applied historically (after adjusting for inflation). The cost estimates include both direct and indirect costs. Note that *average* per person costs are not the same as marginal costs, and therefore the figures used in this analysis cannot be used to calculate the aggregate costs that could be added or saved by increasing or decreasing the total numbers of persons serving sentences.

2.2 Populations of interest

Two study populations were used in the analysis. A sample birth cohort was defined based on people who were born between 1 July 1990 and 30 June 1991 and who were therefore aged 22 as at 30 June 2013. This single 'cohort population' was tracked over time and their risk of poor future outcomes estimated at each year of age from 15 to 22, using the regression modelling approach described below. Characteristics at ages 23 and 24 were not able to be observed for the 1990/91 birth cohort. Information observed at age 22 was used at these ages. Future outcomes were projected beyond age 24 using statistical matching to individuals from an earlier birth cohort, as discussed below.

A second study population was also used, capturing people who were aged 15 to 24 on 31 December 2013, were eligible to live in New Zealand on a permanent basis and were living in New Zealand for at least six months during 2013.⁸ Once the modelling, clustering and target population definition processes were undertaken on the cohort population, the December 2013 (or 'current') population was used to describe the characteristics of the identified target populations. This current population provides a better view of the size and characteristics of at-risk individuals at a recent point in time than the 1990/91 birth cohort population would.

The birth cohort population was based on people born in 1990/91 because the coverage of the various data sets included in the IDI meant their characteristics and outcomes could be tracked up to the age of 22, covering most of the ages of interest in the Youth Funding Review. This was also the first birth cohort for which near-complete school enrolment data was available from the Ministry of Education covering the years when the children were aged

⁷ If more than one sentence was being served simultaneously, the cost estimate applied was that for the highest (most serious) sentence.

⁸ Young people are also excluded if they had no records in the Ministry of Education data or are aged 19 or older and had no records in the Inland Revenue data.

14/15 and above (ie, 2006 and subsequent years). The selection of a cohort based on a 1 July to 30 June year is also consistent with the practice of aligning age to school years.

The criteria for the 1990/91 birth cohort population were intended to select all children who were living in New Zealand as permanent residents during the 2003 to 2007 period, when they were aged 12/13 to 16/17. We selected children who met at least one of the criteria of:

- ▶ being enrolled at a New Zealand school as a domestic student for some or all of the years from 2003 to 2007
- ▶ having an income tax payment record in 2005-08
- ▶ having a benefit paid to them or on their behalf in 2005-07
- ▶ being part of the National Health Index population in 2006-07.

In addition, they had to:

- ▶ be in New Zealand for at least three years of the period from 1 January 2003 to 31 December 2007 (in total, rather than continuously)
- ▶ be born in New Zealand or have permanent residence entitlement through some other means (those with temporary residence visas were excluded).

Defining the birth cohort population in this way has these effects:

- ▶ We miss a small number of children purely because a link could not be established between their administrative data records.
- ▶ We do not include people who were away from New Zealand for much of 2003 and 2007 but were continuously resident at earlier or later phases of their lives.
- ▶ We include some people who were overseas for a substantial part of their childhood or young adulthood. These individuals will be missing from the administrative data sets in earlier and/or subsequent years and will appear to have had no contact with the welfare, child protection or corrections systems. We are able to identify when these people were overseas but do not remove them from the study population.

The second study population comprises children and youth who were aged from 15 to 24 years at 31 December 2013, who had New Zealand citizenship or permanent residence entitlements and were living in New Zealand for at least six months during 2013.⁹

There are 289,540 people aged 15 to 19 and 292,210 people aged 20 to 24 in our current population. These numbers represent 93% and 91% respectively of Statistics New Zealand's estimates of the resident populations in these age groups in the December 2013 quarter. Our study populations are smaller because we exclude temporary residents, those who were out of New Zealand for six months or longer in 2013 and those who could not be linked to the key data sets in the IDI.

⁹ Young people were also excluded if they had no records in the Ministry of Education data or were aged 19 or older and had no records in the Inland Revenue data.

2.3 Estimating future outcomes and costs using statistical matching

A statistical record linkage technique was used to help estimate the likely longer-term outcomes of the study populations. This process is discussed in detail in Treasury Analytical Paper 15/01 in the context of earlier analysis of ICD data and is only summarised briefly in this paper.

The approach involved linking data for an older birth cohort (specifically the July 1978 to June 1979 birth cohort) to the data for the 1990/91 birth cohort population to project outcomes for this latter population. Records were linked on the basis of benefit receipt and corrections sentencing rates and patterns when aged 16 to 21 years inclusive as well as on gender and ethnicity. Observed outcomes and costs experienced by the 1978/79 cohort were then used to estimate the outcomes and costs of the 1990/91 cohort up to age 35.

Matching individuals rather than population groups gives us the flexibility to estimate costs for very different subsets of the population. This is particularly important when we are looking to identify specific target populations for investment decisions. The statistical matching method uses real patterns for individuals over time with very similar observed characteristics up to a certain age.

The approach assumes longitudinal patterns of benefit receipt and corrections sentences can be moved around in time from one cohort to another and that, conditional on a set of 'early indicator' matching variables, these patterns remain relevant to later cohorts. The success of this depends on how well we establish good matching criteria and on how relevant these are for forecasting future outcomes. The range of variables used in the matching process also had some significant omissions, such as region and NCEA achievement. As a result, some caution must be taken with analysis based on these characteristics. Differences in groups defined by these characteristics are probably more diluted than the differences in other group comparisons.

We have also not accounted for differences in macro-economic conditions experienced by the 1978/79 cohort and those that may be faced by the 1990/91 cohort in future years. As a result, future outcome estimates will in part reflect the particular patterns of labour demand and unemployment that have occurred over the last 20 years. Ideally, we would like to remove the effects of these macro-economic fluctuations and have a more constant underlying macro-economic picture underpinning the analysis. This remains an issue for further investigation.

Long-run shifts in New Zealand's social assistance policies could also influence the success of the cohort matching if they have affected the outcomes of different birth cohorts very differently. Ideally, we would adjust individuals' outcomes to remove the effects of any secular trends that are external to the individual but affect the outcomes of the cohort as a whole. In practice, however, it may be difficult to do so in an objective way using the data currently available.

2.4 Approach to the identification of target populations

Identifying risk factors and predicting risk

Logistic regression models were run against the four outcome variables described in section 1.2, covering the welfare, health, education and corrections domains. Over 60 potential risk factors derived from a number of administrative data collections were included in the modelling exercise. Models were run at each year of age for females and males separately. Logistic regression with a forward selection was used to construct a model based on a reduced set of risk factors that were most predictive of each outcome measure. These factors are listed in Appendix 1, along with an indication of the number of models the factor was included in at each year of age.

This process allowed us to identify the key risk predictors for each age/gender combination and calculate an estimated risk score for each individual in the target population. The estimated risk score was used to define an 'at-risk population' according to the above criteria, which could then be used to identify target populations with a higher than average probability of being at risk of poor longer-term outcomes.

As discussed in section 2.3, long-term outcomes were estimated using statistical matching. These were then modelled against characteristics that were directly observed in the data, and this may dilute the relationships between the characteristics and outcomes in the models. Since matching was undertaken on a limited set of characteristics, it is possible that this may not affect all characteristics equally. As such, some caution should be taken when interpreting the relative strength of the modelled relationships.

Defining and describing target populations

For each age group, a cluster analysis was undertaken identifying groups of individuals within the 'at-risk population'. Multiple correspondence analysis was firstly used to redefine the key categorical predictors from the regression modelling into a smaller number of continuous variables, and these were then used to identify a number of clusters at each year of age for females and males jointly.

The youth population was next split into the late teen population (aged 15 to 19) and the early 20s population (aged 20 to 24). Five fairly distinct groups of people with similar characteristics and at particular risk of poor outcomes were identified within each of these age groups. For the early 20s population, risk was defined primarily using the welfare and corrections outcomes measures, as health and education outcomes could have already occurred at these ages, potentially conflating the risk and outcomes measures.

The identification of target population groupings was informed by the factors that were most predictive of poor outcomes in the regression analysis in Step 1 and the clusters identified in Step 2. They were constructed using the following guiding criteria:

- ▶ Parsimony – target populations should be able to be identified using only a few criteria.
- ▶ Separation – overlap between target populations should be minimised.

- ▶ High sensitivity – most people identified as being at risk should fall into at least one target population.
- ▶ High specificity – most people identified as not being at risk should fall outside of the target populations.

2.5 General caveats and cautions

The process of matching records is probabilistic and creates some level of error, as there are likely to be some cases where individuals cannot be matched (and appear in the data with less service use than actually occurred) as well as cases where individuals have been wrongly matched (and appear in the data with inaccurate estimates of service use).

The data covers a specific time and cohort, and some care must be taken in generalising results to the experience of more recent cohorts of children. Some cohorts born more recently have had a higher likelihood of being notified to CYF, partly because of administrative changes related to family violence events attended by Police. This is described in further detail in Treasury Analytical Paper 15/01.

There are also possible biases for those young people who have spent any lengthy period of time outside of New Zealand between ages 15 and 22. The characteristics of these people, including any outcomes achieved, are less likely to be visible in our data, as any contact with government agencies may happen outside of New Zealand. It may look like these people fail to gain qualifications, avoid prison sentences or benefits and do not access health services where these things happen out of New Zealand.

To some degree, this is controlled for by including an indicator in the modelling when a young person is out of the country for the entire previous year. However, there may be some biases introduced that may be better controlled for by including more sophisticated measures of time outside of the country or by treating this group differently, possibly excluding some from the analysis. There is no single approach that would be better, however, and more thinking may be needed on this issue for future work.

3 Which young people experience poor outcomes?

This section presents a descriptive analysis of the 1990/91 birth cohort's characteristics and outcomes observed between ages 15 and 22, alongside their projected outcomes beyond that age. The population is described using a range of factors we expect to be predictive of poor future outcomes. These factors are a selected subset of the factors included in the regression modelling exercise described in the next section. The full set of factors included in the modelling is listed in Appendix 1.

Socio-demographic characteristics

Outcomes by gender and ethnicity are given below in Table 1 for the 1990 cohort population. Young men are somewhat more likely than young women to have poor educational outcomes and considerably more likely to have been sentenced for a criminal offence. Women are slightly more likely to have a poor mental health outcome and considerably more likely to experience long-term benefit receipt in their late 20s and early 30s. Young people of Asian ethnicity are less likely to experience poor outcomes across all domains, while Māori youth have relatively poor outcomes across all but the mental health domain, where they are more or less on a par with European youth. Outcomes for young Pasifika people tend to be better than for Māori but worse than other ethnic groups. The exception to this is mental health service use, where Pasifika rates are low compared to most ethnic groups.

Results are also presented by New Zealand Deprivation Index (NZDep) deciles. NZDep is a geographically defined measure of socio-economic deprivation.¹⁰ Scores are associated with each meshblock in New Zealand and defined in such a way that a 10th of the New Zealand population fall into each decile group. The population living in the least deprived areas in New Zealand are categorised as decile 1, while those living in the most deprived areas are categorised as decile 10. Table 1 shows outcomes for the youth population living in each NZDep decile at age 15. Unsurprisingly, there is a clear gradient of outcomes across NZDep deciles, with those living in higher decile areas being progressively more likely to experience poor outcomes. The exception to this is the mental health outcome measure, with little difference across deciles (ranging from 18% in decile 2 to 22% in decile 8).

¹⁰ See <http://www.otago.ac.nz/wellington/departments/publichealth/research/hirp/otago020194.html> for more information.

Table 1: Socio-demographic characteristics and outcomes for youth 1990/91 cohort

Characteristics	Cohort number	Cohort %	Estimated outcomes					
			No level 2 quals	No level 4 quals	Mental health	Corrections sentence	Long-term benefit	
Gender								
Male	32,118	51%	28%	66%	18%	13%	5%	
Female	30,627	49%	21%	55%	21%	5%	13%	
Ethnicity								
Asian	4,464	7%	12%	42%	8%	2%	2%	
European	39,270	63%	20%	55%	22%	5%	6%	
Māori	13,182	21%	41%	78%	21%	21%	20%	
Other	717	1%	21%	54%	16%	5%	4%	
Pasifika	5,118	8%	30%	71%	11%	11%	11%	
NZDep*								
1 (least deprived)	6,261	10%	13%	47%	19%	4%	4%	
2	6,042	10%	16%	49%	18%	4%	5%	
3	5,886	9%	18%	53%	19%	5%	6%	
4	5,745	9%	19%	55%	20%	6%	6%	
5	5,838	9%	21%	56%	20%	7%	7%	
6	5,883	9%	24%	60%	20%	7%	8%	
7	5,928	9%	27%	63%	21%	9%	10%	
8	6,213	10%	31%	68%	22%	12%	12%	
9	6,912	11%	34%	71%	20%	13%	13%	
10 (most deprived)	8,034	13%	39%	76%	19%	17%	17%	
TOTAL	62,745	100%	25%	60%	20%	9%	9%	

* NZDep is calculated here based on the young person's identified location at age 15.

Location in New Zealand

Outcomes by territorial authority (TA) area are presented in Appendix 2 Table 1. Some caution needs to be exercised when interpreting these results as some areas are small, and results are for a single birth cohort only. Outcomes could vary considerably across cohorts for these small areas. For this reason, we avoid commenting on areas with fewer than 100 people in the cohort. Amongst larger TA areas, a few things stand out however.

A number of areas stand out for the high proportion of people in the cohort at age 15 who were expected to have poor educational outcomes. In Buller, Otago, Ruapehu and Waitomo districts, around two-fifths of 15-year-olds failed to achieve level 2 qualifications by age 23, while around four-fifths failed to achieve a level 4 qualification by age 23. This compares to around one-quarter and three-fifths respectively in the general population and around one-fifth and one-half respectively in Auckland City.

Two territorial authorities stand out not only for poor educational outcomes but for poor outcomes across a range of domains. Kawerau and Wairoa have similar expected educational outcomes to the districts just discussed, but around one-fifth (19% and 22% respectively) of 15-year-olds in the cohort were expected to be sentenced to a custodial or community sentence, and almost one-quarter (23% and 24% respectively) were expected to be on benefit for five years or more between the ages of 25 and 34. The corresponding figures for both measures were about 7% in Auckland City and around 9% across New Zealand overall.

One territorial authority, Carterton, stands out as having particularly high use of mental health services with 34% of the 1990/91 birth cohort using mental health services between ages 20 and 22 (compared to 16% of the Auckland cohort and 20% across New Zealand).

Located overseas

As discussed earlier, where young people spend significant periods of time outside of New Zealand, we may not observe changing characteristics or outcomes in our data. Table 2 looks at this in greater detail by comparing the future outcomes at each age for those in the 1990/91 cohort by whether they were out of New Zealand for the entire year or not. As a result of the way the cohort was defined, limiting the population to people enrolled in schooling around the age of 15 or 16, very few people in our cohort were overseas in the years they turned 15 or 16. Over time, from age 17 progressively, more people moved overseas, with around 4% being overseas for the whole year in the year they turned 22.

Table 2: Outcomes for youth 1990/91 cohort by whether located overseas at each year of age

Location by year of age	Cohort number	Cohort %	Estimated outcomes				
			No level 2 quals	No level 4 quals	Mental health	Corrections sentence	Long-term benefit
Overseas in year turning 15							
No	62,730	100%	25%	60%	20%	9%	9%
Yes	18	0%	0%	67%	33%	0%	0%
Overseas in year turning 16							
No	62,727	100%	25%	60%	20%	9%	9%
Yes	21	0%	43%	71%	0%	0%	0%
Overseas in year turning 17							
No	62,538	100%	24%	60%	20%	9%	9%
Yes	207	0%	84%	94%	7%	7%	6%
Overseas in year turning 18							
No	62,211	99%	24%	60%	20%	9%	9%
Yes	537	1%	84%	93%	4%	7%	5%
Overseas in year turning 19							
No	61,863	99%	24%	60%	20%	9%	9%
Yes	885	1%	75%	96%	3%	5%	4%
Overseas in year turning 20							
No	61,485	98%	24%	60%	20%	9%	9%
Yes	1,263	2%	68%	95%	3%	5%	4%
Overseas in year turning 21							
No	61,128	97%	24%	59%	20%	9%	9%
Yes	1,620	3%	59%	94%	1%	5%	3%
Overseas in year turning 22							
No	60,483	96%	24%	59%	20%	9%	9%
Yes	2,265	4%	53%	88%	3%	5%	3%
TOTAL	62,745	100%	25%	60%	20%	9%	9%

Because of the way the outcomes were constructed, those young people who were overseas in each year show very different outcomes for our different outcome measures of interest. In the case of educational outcomes, lack of contact with the educational system in New Zealand can lead to poor outcomes being inferred when people are not seen to be gaining qualifications. Around 90% or more of those who were overseas for the full year from age 17 through to age 22 did not achieve level 4 qualifications in New Zealand by age 23 compared to around 60% of those who remained in New Zealand. Those who were overseas

at ages 17 or 18 were unlikely to achieve level 2 qualifications in New Zealand (84% did not compared to 24% of those remaining in New Zealand). However, this rate was lower for those who were overseas at older ages (eg, 53% of those who were overseas at age 22).

For other outcomes, extended periods overseas were associated with a reduced likelihood of poor outcomes. In this case, lack of contact with the corrections, welfare or health systems is used to infer the lack of a poor outcome. In particular, very few of those who were overseas after age 15 used mental health services, consistent with the focus of the measure on an earlier age (20 to 22 compared to 25 to 34 for the welfare and corrections outcomes).

Table 2 focuses on people who were overseas for an entire year. Those who were overseas for a substantial part of the year will also have had a reduced likelihood of completing a New Zealand qualification or using benefit, health and corrections services.

Childhood risk factors

By age 15, a number of risk factors may be evident through contact with agencies such as CYF and Work and Income. A number of these potential childhood risk factors are outlined in Table 3 below, alongside future outcomes.

The table shows a clear association of time spent supported by a benefit as a child and future outcomes. There is a clear gradient between the proportion of time spent on a benefit as a child and the likelihood of future poor outcomes across all domains. Those who were mainly supported by a Sole Parent Support benefit (or its equivalent) were particularly likely to experience poor outcomes. This could be due to other risk factors associated with receipt of a Sole Parent Support benefit or could be a reflection of the high likelihood of receipt of such benefits being associated with a considerable length of time on a benefit.

The table also provides information on the parents' or caregivers' educational attainment from benefit data.¹¹ Where a parent has never been on a benefit, no qualification information is collected, and even for those who have been on a benefit, information may not have been collected or may not be up to date. The collection of information could also be connected to other, unobserved characteristics such as the time the caregiver spent on a benefit or the time since they were last on a benefit. As such, the results are difficult to interpret. Nevertheless, conditional on the caregiver having received a benefit and having their qualifications information recorded, higher caregiver qualifications tend to be associated with lower probabilities of poor outcomes. As before, this pattern is less clear with mental health outcomes.

¹¹ Note that the quality of information on educational attainment that is captured in the benefit system is known to be poor.

Table 3: Childhood risk factors and outcomes for youth 1990/91 cohort at age 15

Characteristics at age 15	Cohort number	Cohort %	Estimated outcomes					
			No level 2 quals	No level 4 quals	Mental health	Corrections sentence	Long-term benefit	
Duration supported by benefit as a child								
None	30,636	49%	14%	50%	17%	4%	4%	
1-10%	6,486	10%	21%	59%	19%	6%	7%	
11-25%	4,944	8%	28%	65%	21%	9%	9%	
26-50%	5,961	10%	32%	69%	21%	11%	12%	
50-75%	5,145	8%	38%	75%	24%	14%	15%	
76-85%	1,917	3%	42%	77%	26%	18%	18%	
86-95%	2,055	3%	45%	80%	24%	20%	20%	
95%+	5,607	9%	48%	81%	27%	22%	24%	
Main type of benefit as a child								
None	30,633	49%	14%	50%	17%	4%	4%	
Other	8,610	14%	23%	61%	17%	7%	8%	
Sole Parent Support	23,499	37%	39%	74%	24%	15%	16%	
Maternal caregiver education/benefit status								
Never on benefit	30,636	49%	14%	50%	17%	4%	4%	
On benefit: unknown qualifications	12,294	20%	31%	68%	17%	8%	7%	
On benefit: known qualifications								
No qualifications	12,342	20%	42%	77%	26%	19%	20%	
Level 1 or equivalent	3,021	5%	32%	70%	26%	13%	15%	
Level 2 or equivalent	2,379	4%	27%	64%	24%	12%	13%	
Level 3 equivalent or higher	2,070	3%	23%	58%	25%	10%	11%	
Caregiver with custodial history								
No	36,060	57%	30%	67%	21%	10%	11%	
Yes	3,369	5%	53%	84%	30%	27%	26%	
Unknown	23,316	37%	13%	47%	16%	3%	3%	
Caregiver with community sentence								
No	30,333	48%	27%	64%	20%	9%	9%	
Yes	9,096	14%	48%	82%	28%	23%	22%	
Unknown	23,316	37%	13%	47%	16%	3%	3%	
Notified to CYF care and protection as a child								
No	53,367	85%	21%	57%	18%	6%	6%	
Yes	9,378	15%	48%	80%	32%	23%	24%	
Placed under CYF care and protection								
No	61,644	98%	24%	60%	19%	8%	9%	
Yes	1,104	2%	59%	88%	47%	37%	37%	
CYF care and protection maltreatment finding								
No	58,377	93%	23%	59%	19%	8%	8%	
Yes	4,371	7%	48%	81%	33%	24%	27%	
TOTAL	62,745	100%	25%	60%	20%	9%	9%	

Schooling

Characteristics associated with schooling can also be predictive of future outcomes. Whilst these change between ages 15 and 18, most characteristics do not change after age 18. As such, Table 4 outlines the school-related characteristics of the youth cohort population at age 18 and their projected outcomes.

Around two-thirds of the cohort population were still at school in the year they turned 18. These young people were considerably less likely to be on a benefit long term, to have a corrections sentence or to access mental health services in the future. Not surprisingly, they were much more likely to have gained level 2 qualifications by age 23 and also more likely to have gained level 4 qualifications by age 23.

A small number of young people were overseas at age 18, and for this group, it was unclear whether they were still enrolled at school. While they were unlikely to gain qualifications in the future, they were also unlikely to access mental health services, receive a corrections sentence or be on a benefit, indicating that many probably stay overseas long term and are consequently not captured in the administrative data.

The characteristics of the school the young person most recently attended is also closely associated with a number of outcomes measures. Those young people whose most recent school was a private school or a high decile state-funded school¹² were less likely to experience a range of poor outcomes than those who had attended other schools. As with the earlier analysis of NZDep, the lack of a linear relationship between socio-economic status and mental health service utilisation is reflected in the school decile analysis.

Almost all young people attending a special school failed to achieve level 2 qualifications by age 23 or level 4 qualifications by age 23. Around one-third accessed mental health services, while almost two-thirds were on a benefit long term between ages 25 and 34. Those young people who had accessed special education services experienced similarly high levels of poor outcomes across the education, welfare and mental health domains, while very few received a corrections sentence.

Indicators of behavioural issues at schools are also expected to be associated with poor outcomes, and a history of truancy, suspensions and stand-downs¹³ can be seen to be strongly associated with poor outcomes across all domains. Almost three-quarters of 18-year-olds had achieved level 1 NCEA qualifications by age 18, over one-half had achieved level 2 qualifications and around one-eighth had achieved level 3 qualifications.

¹² State schools and state-integrated schools are funded through the decile system and are allocated a decile score based on the socio-economic characteristics of their school community. See <http://parents.education.govt.nz/primary-school/schooling-in-nz/ministry-funding-deciles/> for more information.

¹³ For more information on stand-downs and suspensions (of which exclusions and expulsions are a subset), see <http://www.educationcounts.govt.nz/indicators/definition/student-engagement-participation/stand-downs-suspensions-exclusions-expulsions>.

Table 4: Schooling characteristics and outcomes for youth 1990/91 cohort at age 18

Characteristics at age 18	Cohort number	Cohort %	Estimated outcomes				
			No level 2 quals	No level 4 quals	Mental health	Corrections sentence	Long-term benefit
Enrolled at school							
No	19,482	31%	53%	80%	27%	18%	16%
Yes	42,729	68%	11%	51%	17%	4%	6%
Unknown	537	1%	83%	93%	4%	6%	4%
Last school decile							
1 (low socio-economic)	3,636	6%	43%	79%	19%	21%	19%
2	4,617	7%	39%	76%	20%	17%	17%
3	4,203	7%	35%	71%	20%	13%	13%
4	6,522	10%	29%	66%	18%	10%	10%
5	7,128	11%	27%	64%	21%	9%	10%
6	6,744	11%	23%	60%	21%	7%	8%
7	7,314	12%	22%	57%	21%	7%	7%
8	5,718	9%	20%	56%	20%	7%	6%
9	5,661	9%	14%	51%	20%	5%	5%
10 (high socio-economic)	7,443	12%	13%	45%	16%	3%	3%
Private school	2,976	5%	12%	46%	18%	2%	2%
Currently in special school							
No	62,574	100%	24%	60%	20%	9%	9%
Yes	174	0%	90%	97%	34%	10%	62%
Ever truant from school							
No	58,431	93%	22%	59%	19%	7%	8%
Yes	4,317	7%	61%	86%	30%	26%	25%
Ever suspended from school							
No	59,718	95%	23%	59%	19%	8%	8%
Yes	3,027	5%	55%	86%	34%	30%	20%
Ever stood down from school							
No	54,075	86%	21%	57%	18%	7%	8%
Yes	8,673	14%	48%	81%	30%	22%	17%
Ever received special education services							
No	62,352	99%	24%	60%	20%	9%	9%
Yes	399	1%	85%	93%	28%	2%	66%
Achieved NCEA level 1 or equivalent							
No	18,342	29%	64%	84%	27%	19%	19%
Yes	44,403	71%	8%	51%	17%	5%	5%
Achieved NCEA level 2 or equivalent							
No	27,663	44%	56%	80%	25%	15%	15%
Yes	35,088	56%	n/a	45%	16%	4%	4%
Achieved NCEA level 3 or equivalent							
No	53,691	86%	29%	65%	21%	10%	10%
Yes	9,060	14%	n/a	34%	14%	4%	4%
TOTAL	62,745	100%	25%	60%	20%	9%	9%

Tertiary education

Tertiary qualifications are clearly likely to be related to educational outcomes, but tertiary study is also expected to be linked with broader social outcomes. Table 5 shows participation and outcomes of tertiary study¹⁴ by age 22 alongside the broader set of outcomes analysed.

Table 5: Tertiary education and outcomes for youth 1990/91 cohort at age 22

Characteristics at age 22	Cohort number	Cohort %	Estimated outcomes				
			No level 2 quals	No level 4 quals	Mental health	Corrections sentence	Long-term benefit
Highest qualification level							
0	12,447	20%	90%	95%	28%	20%	21%
1	4,818	8%	89%	95%	22%	11%	12%
2	10,089	16%	0%	90%	20%	8%	7%
3	19,002	30%	0%	66%	17%	5%	5%
4	6,621	11%	0%	0%	18%	6%	7%
5	2,352	4%	0%	0%	17%	4%	5%
6	1,125	2%	0%	0%	15%	3%	3%
7	5,697	9%	0%	0%	12%	1%	2%
8	600	1%	0%	0%	10%	2%	0%
Highest tertiary level enrolment							
None	16,857	27%	57%	94%	20%	13%	12%
1	408	1%	74%	97%	26%	18%	35%
2	3,090	5%	38%	92%	27%	19%	16%
3	6,396	10%	27%	92%	24%	15%	16%
4	7,056	11%	19%	49%	23%	12%	13%
5	3,498	6%	11%	41%	20%	7%	9%
6	1,980	3%	8%	45%	18%	5%	5%
7	19,740	31%	3%	33%	17%	2%	3%
8	3,723	6%	2%	17%	13%	1%	1%
1 year tertiary study							
No	25,479	41%	45%	88%	22%	13%	12%
Yes	37,266	59%	11%	42%	18%	6%	7%
2 years tertiary study							
No	37,041	59%	38%	81%	22%	12%	12%
Yes	25,707	41%	6%	31%	17%	4%	5%
3 years tertiary study							
No	46,638	74%	32%	73%	21%	11%	11%
Yes	16,110	26%	4%	24%	15%	3%	4%
TOTAL	62,745	100%	25%	60%	20%	9%	9%

One-fifth of the 22-year-old youth population had not achieved a qualification by age 22, while around one-tenth had only achieved a level 1 qualification. Not surprisingly, almost all of these groups failed to achieve a level 2 or 4 qualification by age 23. By construction, nobody who had achieved a level 2 or 3 qualification (around half of the cohort) at age 22 could have failed to achieve this level by age 23, and similarly, none of the approximately one-quarter of young people who had achieved level 4 to 8 qualification at age 22 were recorded as having a poor educational outcome under either measure.

¹⁴ In the case of highest qualification level, this also includes school qualifications.

In general, high qualifications and high-level tertiary qualification enrolments are associated with good outcomes across most domains. This is most clearly seen in corrections sentencing with 5% or fewer of those with level 5 qualifications or above at age 23 having a future corrections sentence or long-term benefit receipt between ages 25 and 34. Mental health outcomes show a less strong declining trend with higher qualifications. However, 28% of those with no qualifications at age 22 used mental health services between ages 20 and 22, whilst only 10% of those with level 8 qualifications did so. In the case of tertiary enrolments, people who had been enrolled in level 1 or 2 tertiary qualifications had the highest rates of mental health service use, at 26% and 27% respectively. Those enrolled in level 1 tertiary qualifications were most likely to be on a benefit long term, whilst those enrolled in level 2 tertiary qualifications were most likely to receive a corrections sentence. Enrolment in low-level tertiary qualifications could be a proxy for early school leaving and a failure to achieve this level of qualification at school.

The longer someone is in tertiary study before age 22, the less likely they are to experience poor future outcomes across all domains. Only one-quarter of those with three years' tertiary study by age 22 failed to achieve a level 4 qualification by age 23, while 15% used mental health services between ages 20 and 22, only 3% received a corrections sentence between ages 25 and 34 and only 4% received a benefit long term in the same age range.

Work and welfare

We would expect employment and earnings to be important predictors of future outcomes across multiple domains. In addition, both time on a benefit and type of benefit are likely to be closely linked with multiple outcome domains. For example, receipt of the Supported Living Payment – Health Condition or Disability may be strongly associated with outcomes across the health, welfare and education domains. Table 6 below presents selected measures of employment, earnings and benefit receipt, by future outcomes, at age 22.

At age 22, almost one-quarter of young people in the 1990/91 birth cohort had no salary, wage or self-employment earnings.¹⁵ As might be expected, this group had poor future outcomes across all domains, with three-quarters not achieving level 4 qualifications by age 23 and one-fifth expected to be on a benefit for five years or more between ages 25 and 34. In general, progressively higher earnings were associated with improved outcomes. Of those earning over \$40,000 at age 22, only 2% were expected to be on a benefit long term, 6% were expected to receive a corrections sentence and 14% were expected to use mental health services.

On the other hand, two-thirds of those earning over \$30,000 at age 22 failed to achieve level 4 qualifications by age 23. This is perhaps not surprising given that this level of earnings is likely to be strongly associated with full-time work that would preclude full-time study. Lower earnings are consistent with part-time work that, in many cases, would be undertaken alongside full-time study.

¹⁵ This includes 4% who were overseas for the entire year and others who were overseas for part of the year.

Table 6: Employment and welfare characteristics and outcomes for youth 1990/91 cohort at age 22

Characteristics at age 22	Cohort number	Cohort %	Estimated outcomes					
			No level 2 quals	No level 4 quals	Mental health	Corrections sentence	Long-term benefit	
Earnings last year								
None	14,931	24%	42%	75%	22%	13%	20%	
\$5,000 or less	9,375	15%	22%	52%	23%	11%	12%	
\$5,001 to \$7,500	3,324	5%	17%	45%	20%	8%	7%	
\$7,501 to \$10,000	3,135	5%	15%	42%	18%	7%	7%	
\$10,001 to \$15,000	5,343	9%	15%	45%	19%	6%	5%	
\$15,001 to \$20,000	4,395	7%	17%	50%	20%	7%	5%	
\$20,001 to \$30,000	7,470	12%	20%	61%	19%	7%	4%	
\$30,001 to \$40,000	7,761	12%	22%	66%	17%	6%	3%	
over \$40,000	7,011	11%	20%	66%	14%	6%	2%	
Proportion of time NEET since age 16								
None	8,397	13%	12%	51%	15%	3%	2%	
1-10%	25,356	40%	14%	49%	15%	3%	3%	
11-25%	13,134	21%	21%	59%	19%	7%	6%	
26-50%	8,313	13%	39%	78%	27%	15%	17%	
50-75%	4,953	8%	57%	88%	34%	26%	31%	
76-85%	1,329	2%	75%	94%	40%	33%	40%	
86-95%	873	1%	80%	97%	38%	37%	42%	
95%+	402	1%	93%	99%	37%	31%	43%	
Proportion of time on a benefit since age 18								
None	33,540	53%	16%	53%	13%	3%	2%	
1-10%	8,661	14%	17%	53%	19%	6%	3%	
11-25%	7,038	11%	26%	63%	23%	10%	6%	
26-50%	5,343	9%	40%	77%	30%	20%	16%	
50-75%	3,369	5%	49%	83%	36%	26%	31%	
76-85%	1,146	2%	59%	87%	40%	26%	43%	
86-95%	1,047	2%	59%	88%	44%	29%	50%	
95%+	2,607	4%	67%	89%	40%	20%	58%	
Main benefit type since age 18								
None	33,540	53%	16%	53%	13%	3%	1%	
Youth Payment	630	1%	39%	76%	30%	16%	10%	
Young Parent Payment	117	0%	67%	92%	31%	28%	41%	
Jobseeker – Health and Disability	3,441	5%	43%	79%	59%	19%	25%	
Jobseeker – Work Ready	13,401	21%	36%	75%	24%	17%	11%	
Jobseeker – Training	6,456	10%	11%	39%	16%	6%	6%	
Supported Living Payment – Carer	189	0%	38%	73%	17%	17%	22%	
Supported Living Payment – Health	1,170	2%	67%	86%	42%	9%	62%	
Sole Parent Support	3,798	6%	50%	83%	27%	17%	40%	
TOTAL	62,745	100%	25%	60%	20%	9%	9%	

An alternative measure of engagement that combines both a labour market and education dimension is the proportion of time spent not in employment, education or training (NEET) since age 16. Around one-half of the youth population spent either no time or less than 10% of time between ages 16 and 22 as NEET. Both of these groups had relatively good

outcomes across all domains. Another one-fifth spent between 10% and 25% of their time as NEET, while around one-quarter spent more than 25% of their time as NEET.

As might be expected, there is a clear relationship between time spent NEET and the expectation of poor outcomes across all domains. Of the 4% of young people who had spent three-quarters or more of their time as NEET between ages 16 and 22, almost 40% had accessed mental health services by age 25, more than one-third were sentenced to a custodial or community sentence between ages 25 and 34 and over 40% were on a benefit for more than five years between the same ages. Not surprisingly, almost all this group failed to achieve level 2 and level 4 qualifications by ages 18 and 23 respectively.

Benefit receipt was also closely associated with outcomes. Not surprisingly, progressively more time on a benefit was particularly strongly associated with future time in receipt of a benefit, but there were also strong associations with other outcomes. Most benefit types were also strongly associated with poor outcomes across multiple domains. People who had mainly spent time on a Jobseeker Support – Training benefit at age 22 (10% of all young people) had a lower probability of poor educational outcomes than other beneficiaries and even than those who had not spent time on a benefit. They also had considerably better outcomes across other domains than those on other benefits.

Not surprisingly, those 22-year-olds whose main benefit since age 18 was either Jobseeker Support – Health and Disability or Supported Living Payment – Health and Disability were more likely than other groups to have used mental health services between ages 20 and 22, but the latter groups also had relatively poor educational outcomes and were expected to be on a benefit long term between ages 25 and 34. While they were very few in number, those who had mainly been on a Young Parent Payment benefit at age 22 had a high likelihood of poor outcomes across all domains and were particularly likely to receive a corrections sentence between ages 25 and 34.

Early corrections contact

Early contact with the corrections system, either through contact with CYF Youth Justice or through receiving a custodial or community sentence as an adult, would be expected to be closely linked with the chances of being sentenced to a custodial or community sentence between ages 25 and 34. However, these may also be important predictors of outcomes across other domains. Table 7 shows expected outcomes by contact with the corrections system up to age 22.

Around 5% of young people had been referred to CYF Youth Justice by age 22. Not surprisingly, this group were particularly likely to receive a corrections sentence between ages 25 and 34 (almost half compared to 7% of those who had not been referred to Youth Justice by age 22). Of the 1% of youth who had a CYF Youth Justice placement by age 22, around two-thirds received a community or custodial sentence between 25 and 34. Both of these groups (especially those with a placement) also had poor outcomes across other domains with large numbers having failed to achieve level 2 and level 4 qualifications by ages 18 and 23 respectively and large numbers also being users of mental health services (41% and 58% of those with a referral and a placement respectively).

Almost one-tenth of young people had received a community sentence by age 22, while around 3% had received a custodial sentence. Not surprisingly, being sentenced by age 22 was highly associated with being sentenced between ages 25 and 34, with longer sentences progressively increasing this probability. Over 80% of those young people who had spent a year or more on a custodial sentence by age 22 received a corrections sentence between ages 25 and 34.

As with youth justice contact, being sentenced at an early age was associated with poor outcomes across all domains, although length of sentence (particularly for custodial sentences) did not have as strong a relationship with other outcomes as it did for the chances of reoffending.

Table 7: Corrections contact and outcomes for youth 1990/91 cohort at age 22

Characteristics at age 22	Cohort number	Cohort %	Estimated outcomes					
			No level 2 quals	No level 4 quals	Mental health	Corrections sentence	Long-term benefit	
CYF Youth Justice referral								
No	59,310	95%	23%	59%	18%	7%	8%	
Yes	3,438	5%	62%	88%	41%	46%	26%	
CYF Youth Justice placement								
No	62,364	99%	24%	60%	19%	8%	9%	
Yes	381	1%	71%	94%	58%	67%	35%	
Community sentence length								
None	56,844	91%	21%	58%	17%	5%	8%	
6 months or less	1,779	3%	51%	84%	31%	40%	20%	
6 months to 1 year	1,677	3%	55%	87%	48%	43%	22%	
1 year to 2 years	1,605	3%	62%	92%	53%	57%	25%	
2 years plus	843	1%	68%	90%	58%	59%	32%	
Custodial sentence length								
None	60,948	97%	24%	59%	18%	7%	8%	
6 months or less	924	1%	68%	92%	61%	62%	30%	
6 months to 1 year	432	1%	61%	89%	60%	64%	26%	
1 year to 2 years	246	0%	60%	95%	66%	82%	32%	
2 years plus	198	0%	56%	97%	59%	85%	29%	
TOTAL	62,745	100%	25%	60%	20%	9%	9%	

Early use of mental health services

Early indicators of mental health issues could be a strong predictor of future outcomes across multiple domains, with mental health issues potentially affecting participation in education and the labour market. Mental health issues, and especially alcohol and drug addiction, are also potentially linked with criminal offending in various ways. Table 8 shows four indicators of poor mental health at age 20 by future outcomes.

One in 20 young New Zealanders who were born in the year to the end of June 1991 had used alcohol or drug addiction services by age 20, while almost 10% had used some other

sort of mental health services.¹⁶ In both cases, this was associated with poor educational, welfare and mental health outcomes. Not surprisingly, more than three-quarters of young people who had accessed these services by age 20 continued to access them between ages 20 and 22. Almost one-quarter were on a benefit long term between ages 25 and 34 (compared to fewer than one-tenth of other young New Zealanders). Young people who had accessed alcohol or drug addiction services by age 20 were particularly likely to serve a future corrections sentence, with 40% doing so between the ages of 25 and 34 compared to 7% of those who hadn't accessed these services.

Broader indicators of substance abuse and other mental health issues were able to be derived using a wider set of health data including prescriptions for mental health-related pharmaceuticals. Around one-tenth of young people were identified as having a history of substance abuse and one-quarter of having mental health issues according to these expanded measures. Both measures show a relationship with all outcome domains, although in most cases, this was more muted than the earlier measures.

Table 8: Use of mental health services for youth 1990/91 cohort at age 20

Characteristics at age 20	Cohort number	Cohort %	Estimated outcomes					
			No level 2 quals	No level 4 quals	Mental health	Corrections sentence	Long-term benefit	
Used alcohol or drug addiction services								
No	61,197	98%	24%	60%	19%	8%	9%	
Yes	1,551	2%	53%	84%	59%	40%	23%	
Used other mental health services								
No	58,959	94%	23%	59%	17%	8%	8%	
Yes	3,789	6%	43%	77%	63%	22%	25%	
Indicator of substance abuse								
No	58,617	93%	24%	60%	19%	8%	9%	
Yes	4,131	7%	34%	67%	35%	14%	16%	
Indicator of other mental illness								
No	50,607	81%	22%	58%	13%	7%	7%	
Yes	12,141	19%	36%	70%	47%	16%	17%	
TOTAL	62,745	100%	25%	60%	20%	9%	9%	

Early parenting

Early parenting could be associated with poor outcomes for a number of reasons. In particular, childcare could affect the ability to participate in employment or education. Table 9 shows outcomes by early parenting status at age 22, whether the young person was a parent before age 19 and the nature of any interaction with CYF with regard to that child or children.

In total, 13% of the 1990/91 birth cohort had had a child or children by age 22, while 5% were a parent before age 19. Becoming a parent at an early age was associated with a high

¹⁶ Due to the fact that data on mental health and alcohol or drug addiction service use is not available until 2008 when the youth in this cohort were aged 16/17, these figures are likely to be underestimates of the proportion of people who had used these services.

likelihood of poor outcomes across all domains, with those who became a parent by 19 being slightly more at risk than those who had a child for the first time in the following three years.

Where a young parent had a child that was subject to a CYF care and protection notification (4% of the birth cohort), outcomes were particularly poor. Almost one-third of these young people had a corrections sentence between ages 25 and 34, while almost two-fifths were on a benefit long term during this period, and a similar number accessed mental health services between ages 20 and 22. Where there was a finding of abuse or where there had been a Police/family violence notification, these figures were even higher, and for the 228 parents whose children had been placed in care by age 22, outcomes were particularly poor. Two-fifths of this group had a poor corrections outcome, over one-half were on a benefit long term and almost three-fifths used mental health services. Two-thirds had not achieved NCEA level 2.

Table 9: Early parenting and offspring childhood risk factors for youth 1990/91 cohort at age 22

Characteristics at age 22	Cohort number	Cohort %	Estimated outcomes					
			No level 2 quals	No level 4 quals	Mental health	Corrections sentence	Long-term benefit	
Parent								
No	54,465	87%	21%	57%	19%	7%	6%	
Yes	8,280	13%	49%	84%	27%	19%	26%	
Parent before 19								
No	59,808	95%	23%	59%	19%	8%	8%	
Yes	2,937	5%	56%	86%	29%	22%	32%	
Offspring CYF care and protection notification								
No	60,288	96%	23%	59%	19%	8%	8%	
Yes	2,460	4%	63%	90%	42%	31%	39%	
Offspring any findings of abuse								
No	61,419	98%	24%	60%	19%	8%	8%	
Yes	1,329	2%	64%	90%	45%	36%	43%	
Offspring CYF care and protection placement								
No	62,520	100%	25%	60%	20%	9%	9%	
Yes	228	0%	67%	88%	58%	41%	53%	
Offspring Police/family violence notification								
No	61,194	98%	24%	60%	19%	8%	8%	
Yes	1,554	2%	63%	88%	41%	34%	42%	
TOTAL	62,745	100%	25%	60%	20%	9%	9%	

4 Predicting poor outcomes

In this section, regression analysis is used to more systematically identify which characteristics observed in the administrative data between ages 15 and 24 were most strongly associated with a higher likelihood of poor outcomes as an adult. The aims of this analysis are predictive in nature, seeking to identify the factors most strongly associated with poor outcomes and to use these to predict risk and identify target populations for investment decision making. The analysis does not seek to understand the causes of poor outcomes and does not answer underlying questions of causality. Just because a factor predicts poor outcomes does not necessarily mean the underlying concept causes those outcomes.

We are restricted to using existing administrative data sources from each agency. We are also limited by the confines of the data collected in the IDI for the cohort we are studying. This meant, for example, that health data relating to early childhood was not able to be used in the study. As such, the results are less definitive in terms of the relative importance of various factors. Nevertheless, the analysis is useful for understanding which interactions with government agencies at particular stages of a young person's life are more strongly associated with poor outcomes later on.

As discussed above in section 2, logistic regression models were run for the 1990/91 birth cohort for each year of age from the year they turned 15 to the year they turned 22 (ie, from the 2005/06 July to June year to the 2012/13 year). Predictive factors were selected for the models on the basis of a forward selection approach.¹⁷ Models were also run separately for males and females on the basis that different risk factors were likely to be important for each gender. Models were run separately for our four outcome measures, with models on the health and educational outcomes being run up to ages 20 and 21 respectively to avoid the predictive factors becoming conflated with the outcome measures. These outcome measures are measured at an earlier age than the welfare and corrections measures.

4.1 Regression model factor selection and estimation

Predictive factors selected in the models

Appendix 1 Table 1 highlights the number of times each characteristic is selected across the different models at each age. This gives a broad idea of the characteristics that are important in predicting risk as someone ages through the late teen years and into their 20s. Care needs to be taken in interpreting the importance of these selections however.

The choice of factors to include in a forward selection modelling procedure is heavily dependent on the factors already selected in the model. Where a factor is highly correlated with another factor already included, it may not add much to the model and hence not be selected for the final model. With very slightly different data, the reverse may be true. For example, duration spent on a benefit is closely associated with the type of benefit, and each

¹⁷ Appendix 1 summarises the number of models each factor was included in at different ages (out of a possible eight in each year up to age 20, six at age 21 and four at age 22).

may be related to future time on a benefit. In cases where the duration is slightly more predictive of future benefit receipt and hence added to the model first, benefit type may not be included, even though it is also predictive of future receipt. With slightly different data (and possibly depending on what other variables are already added to the model, for example, use of mental health services or early parenting status), benefit type may be included but not duration.

Nevertheless, there are some interesting patterns in the risk factors selected for the models. Broadly speaking, as people age from 15 to 22, we have more information about them that can be used to predict future outcomes. At age 15, there were 42 potential factors used in the modelling, while by age 18, there were more than 60. With an increase in the number of potential factors, more factors were generally selected for the models. On average, 15.6 factors were used per model at age 15, increasing to around 21 at ages 18 to 20. While fewer factors were used in the models at ages 21 and (especially) 22, these were ages at which only welfare and corrections outcomes were being predicted, with fewer models run as a result.

Some specific patterns are evident in the table and are worth pointing out:

- ▶ Some factors are clearly predictive across all outcomes and most ages. The most prominent of these is ethnicity, which is included in all 58 models, but 'Notified to CYF care and protection as a child' was included in 56 models (and all models up to age 20), 'Maternal caregiver education/benefit status' was included in 53 models (and all models up to age 19), and 'Referred to youth justice' was included in 48 models.
- ▶ The only factors not included in any model were the 'Early parent (before age 19)' indicator and the 'Had own child in placement or with maltreatment finding' indicator. The former may be highly correlated with some benefit types, while the latter is closely linked to other indicators regarding interactions around the young person's child(ren), many of which were included in a few models.
- ▶ As might be expected, characteristics relating to school-level qualifications were mainly important during the mid to late teenage years. The NCEA level 1 achievement indicator was used in all models at age 16 but no models after age 18, whilst levels 2 and 3 were most important at ages 18 to 20. Having been stood down from school was a significant factor for most models at ages 15 to 20, having been suspended from school was important at ages 15 to 16 and being recorded as being truant from school was important to most outcomes at age 17. Having received special education services was predictive in at least half of the models at all ages. School decile was important in most models at ages 15 to 17.
- ▶ A number of factors were constructed relating to the enrolment and completion of tertiary qualifications, and these measures were included in various models from ages 18 to 21.
- ▶ Simple yes or no indicators of employment were included in five of the models at ages 15 and 16, but the level of earnings became more important as a predictor of outcomes by the later teen years. Depending on the model, the factor selected related to the previous year or the previous two years. However, the variables are highly correlated, and the distinction may not be meaningful. Time spent NEET was important from ages 17 to 21

(not being available prior to age 17), with different factors constructed that covered different time periods. Indicators of benefit status, type and duration were included in all models from age 18, the minimum age of eligibility for most types of benefit.

- ▶ Factors related to the young person's caregiver were particularly important at the younger ages. Having a caregiver with a community sentence was included in almost all models through to age 18, while having a caregiver with a custodial sentence was included in half of the models at age 15.
- ▶ Unsurprisingly, accessing mental health services or being sentenced to a community or custodial sentence at any age were important predictors of poor mental health and corrections outcomes respectively. However, they were also broadly predictive of poor outcomes across other domains. In the case of corrections sentences, whether the sentence was custodial or community appears to be of limited importance in predicting outcomes. However, accessing alcohol or drug services appears to predict outcomes quite differently from accessing other mental health services, with the latter being much more broadly predictive across multiple outcomes domains.

Model discrimination

The area under the receiver operating characteristic (ROC) curve indicates how well each model is able to differentiate between those young people at each age who go on to have poor outcomes as adults and those that do not. The ROC statistic is a measure of how well a logistic regression model fits the data. Specifically, it measures how well the model discriminates between those with and without the outcome of interest.

The areas under the ROC curves for each of the 54 models that were run are given in Table 10 below. The model that fitted least well was that predicting future mental health outcomes for 15-year-old females (ROC statistic of 0.64), while the models that fitted the best were generally those predicting a corrections sentence or longterm benefit receipt at ages 20 to 22 (ROC statistics consistently above 0.8). The average across all 54 models was 0.80, indicating that the models were generally good at predicting who would experience a poor future outcome.

Comparing females to males, there was little difference in the ROC statistic, with the models for females having slightly better fit in general but only marginally so. Consistent with both more information becoming available over time (often closely linked to the outcomes of interest) and increasing proximity to the outcome period, predictions generally improved as a person aged. Average ROC statistics increased from around 0.75 at age 15 to almost 0.9 at age 22.

Some future outcomes also appear to be easier to predict at an early age than others. Averaged across ages 15 to 19 (ages 20 to 22 are excluded since not all outcomes were modelled), corrections and welfare outcomes had higher ROC statistics than the other two outcomes on average and at each year of age. Across all ages, the use of mental health services was clearly the most difficult to predict, with ROC scores considerably smaller than for other outcomes. This is perhaps not surprising given the earlier descriptive analysis, which showed less clear differentiation in mental health outcomes across key socio-demographic characteristics such as ethnicity, deprivation decile and school decile.

High ROC statistics at the older ages (especially 19 years and over) reflect the availability of measures that are closely related to the outcomes being modelled (for example, benefit receipt), as well as the close proximity of the age at which outcomes are measured. At age 19, for example, it is relatively easy to predict whether somebody will achieve a level 4 qualification by age 23, as qualifications achieved up to age 19 are known, as is the level of any current study being undertaken at that age.

Table 10: Areas under receiver operating characteristic (ROC) curves for each youth outcome model

Age	Model by outcome				Average
	No Level 2/4 Quals *	Mental Health	Corrections sentence	Longterm benefit	
Female					
15	0.77	0.64	0.82	0.80	0.76
16	0.80	0.66	0.83	0.81	0.78
17	0.75	0.68	0.85	0.83	0.78
18	0.79	0.70	0.85	0.84	0.80
19	0.85	0.72	0.87	0.86	0.83
20	n/a	0.76	0.88	0.87	0.84
21	n/a	n/a	0.89	0.88	0.88
22	n/a	n/a	0.89	0.89	0.89
Average 15-19	0.79	0.68	0.85	0.83	0.79
Male					
15	0.74	0.66	0.78	0.77	0.74
16	0.77	0.68	0.79	0.79	0.76
17	0.74	0.69	0.81	0.81	0.76
18	0.77	0.70	0.82	0.83	0.78
19	0.83	0.72	0.83	0.86	0.81
20	n/a	0.75	0.84	0.88	0.84
21	n/a	n/a	0.85	0.89	0.86
22	n/a	n/a	0.86	0.89	0.88
Average 15-19	0.77	0.69	0.81	0.81	0.77
Average ALL	0.78	0.70	0.84	0.84	0.80

* Level 2 qualifications were modelled at ages 15 and 16, and level 4 at older ages.

4.2 Predicting risk and defining populations with high risk

The regression modelling allowed us to construct an equation for each individual that could be used to allocate them a risk score for each outcome of interest based on their age and gender as well as a wide range of other characteristics. Individual data is anonymised, and as such, it is difficult to use an individual risk score to target services. For this reason, the main purpose of this study is to identify target populations with high risk of poor outcomes across our outcome domains based on a small set of identifiable characteristics. The first stage in getting to these target populations was to construct a measure or measures of broader risk. We constructed two measures that were used in the remaining analysis:

- ▶ Risk across multiple poor outcomes – at each year of age, the population was ordered according to their estimated risk score for each outcome and assigned a rank. These ranks were then averaged, and the population was ordered according to this average rank. Following a fairly arbitrary delineation, the 5% of the population with the highest average ranks were defined as being at **extreme risk**, while the 10% with the next highest average ranks were defined as being at **high risk**.¹⁸
- ▶ Extreme risk of at least one poor outcome – the ranks constructed in the previous process were used to identify the 5% of the population at greatest risk of a poor outcome on each outcome measure, ie, at extreme risk of that outcome. A person was considered to be **at risk** where they were at extreme risk for at least one of the four outcomes.

The process of calculating risk scores and ranks and identifying general risk measures was repeated for both the 1990/91 birth cohort population and the December 2013 population. The focus of the descriptive analysis in the remainder of the report uses the December 2013 population as its basis, although outcomes measures and costs are inferred from the equivalent 1990/91 population either according to level of risk or target populations, as defined in the next section.

Risk across multiple poor outcomes – characteristics and expected outcomes

Table 11 below shows the different demographic characteristics of the current population according to different levels of estimated risk across multiple outcomes as defined above. Almost two-thirds of the youth population defined as being at extreme risk are Māori, as are a little over half of the high-risk population. This compares to 13% of the remaining youth population defined as having low to moderate risk. A little over half of the extreme-risk population are male, while almost half live in low socio-economic status areas, as defined by small area deprivation (NZDep) deciles 9 and 10. Only 20% of the low to moderate-risk population live in these areas.

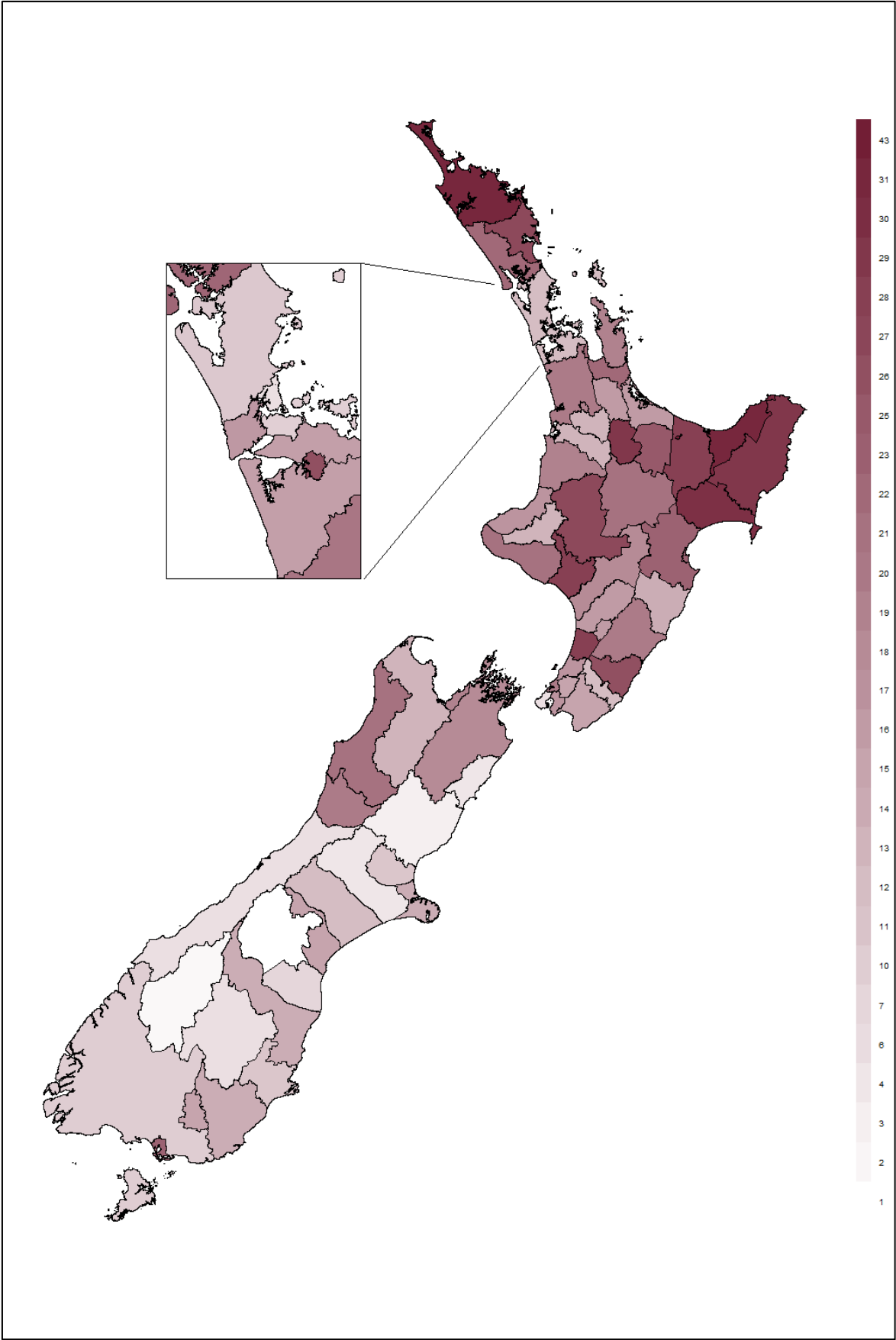
Table 11: Demographic characteristics of the December 2013 youth population by level of risk across multiple outcomes

	Total 2013 youth population	Extreme risk (5% most at risk)	High risk (10% next at risk)	Everyone else (85% least at risk)
Number	581,740	29,080	58,170	494,490
% Male	51%	55%	53%	51%
% Māori	19%	62%	50%	13%
% European/Pākehā	61%	31%	39%	65%
% Pasifika	9%	7%	10%	9%
% Low SES (NZDep 9 and 10)	24%	49%	43%	20%

The distribution of the high-risk population across New Zealand territorial authority areas is given in Figure 1. The numbers behind this figure are presented in Appendix 3 Table 1.

¹⁸ Note that the high-risk population generally refers to the population meeting at least the definition of high risk, and includes those identified as being at extreme risk.

Figure 1: Percentage of the December 2013 youth population at high risk across multiple poor outcomes by territorial authority area (Auckland territorial authorities expanded)



The top four regions in terms of this youth risk measure are Kawerau, Opotiki, Far North and Wairoa districts (with 42%, 31%, 30% and 30% high risk respectively). These areas have relatively small youth populations, however, and as such, only 4% of high-risk young people live in these areas. This compares to large urban areas such as Manukau City, where almost 10% of all high-risk young people live, and Waitakere City, where 5% live.

Not surprisingly, given the modelling was designed to predict poor future educational, welfare, mental health and corrections outcomes, high-risk groups had considerably poorer expected future outcomes and higher projected future welfare and corrections costs. Table 12 shows these projected outcomes by estimated overall risk.

Table 12 shows a clear pattern of increasing risk of poor outcomes with increasing estimated overall risk (for example, almost all of the extreme-risk population fail to achieve NCEA level 4 by age 23 compared to half of the low to moderate-risk population). Risk is nevertheless difficult to predict. While relatively high proportions of the high-risk population are expected to go on to be supported by a benefit long term between ages 25 and 34 (26% and 40% respectively compared to 5% of the lower risk population), this only represents a little over half of the 57,000 people expected to be long term on a benefit. Even low proportions of poor outcomes amongst the large low to moderate-risk population can equate to large numbers of people.

Table 12: Estimated outcomes for the 1990/91 birth cohort population by level of risk across multiple outcomes

	1990/91 birth cohort population	Extreme risk (5% most at risk)	High risk (10% next at risk)	Everyone else (85% least at risk)
Benefit 5+ years	9.0%	39.5%	26.1%	5.2%
Corrections sentence	8.7%	45.7%	23.9%	4.8%
No level 2 by 23	24.7%	71.8%	58.2%	17.9%
No level 4 by 23	60.4%	95.2%	90.2%	54.8%
Mental health services	19.7%	58.4%	35.7%	15.5%
Projected corrections and benefit costs age 25-34	\$28,000	\$131,000	\$74,400	\$16,600

Table 13: Projected welfare and corrections costs aged 25 to 34 by estimated risk across multiple outcomes at age 20, 1990/91 birth cohort

	1990/91 birth cohort population at age 20	Extreme risk (5% most at risk)	High risk (10% next at risk)	Everyone else (85% least at risk)
\$0	35,004	537	1,992	32,475
\$1 to \$100,000	20,886	1,245	2,652	16,989
\$100,001 to \$200,000	4,215	717	972	2,526
\$200,001+	2,376	621	621	1,134
Total	62,481	3,120	6,237	53,124

Similarly, while expected welfare and corrections costs for the extreme-risk population are almost 10 times higher than those of the low to moderate-risk population, large numbers of those estimated to have low to moderate risk are estimated to have high future welfare and

corrections costs. Table 13 shows estimated future costs in broad dollar bands by level of estimated risk for the 1990/91 birth cohort population at age 20. Whilst only a small proportion of young people with predicted high or extreme risk are expected to have zero future corrections and benefit costs, reasonably large numbers (albeit a small percentage) of those with low to moderate risk are expected to have future costs in excess of \$200,000.

Extreme risk of one or more outcomes – characteristics and expected outcomes

Tables 14 and 15 below show the demographic characteristics and expected outcomes respectively where the population is divided according to whether a person is at extreme risk of one or more outcomes or not. Around half of the at-risk youth population are Māori compared to 15% of the remaining youth population.

As expected, the at-risk population has poorer expected outcomes than the not-at-risk population across all outcome measures. When compared to the high and extreme-risk populations identified above, the at-risk population expected outcomes generally lie between the two, as we might expect (ie, poor outcomes are more likely than for the high-risk population but less likely than for the extreme-risk population). The exception to this is the 'No level 4 qualifications by age 23' outcome for which the at-risk population have slightly better outcomes than either the high-risk or extreme-risk populations (88% compared to 90% and 95% respectively). It's possible that this outcome is less predictive of extreme risk than it is of lower levels of broader risk.

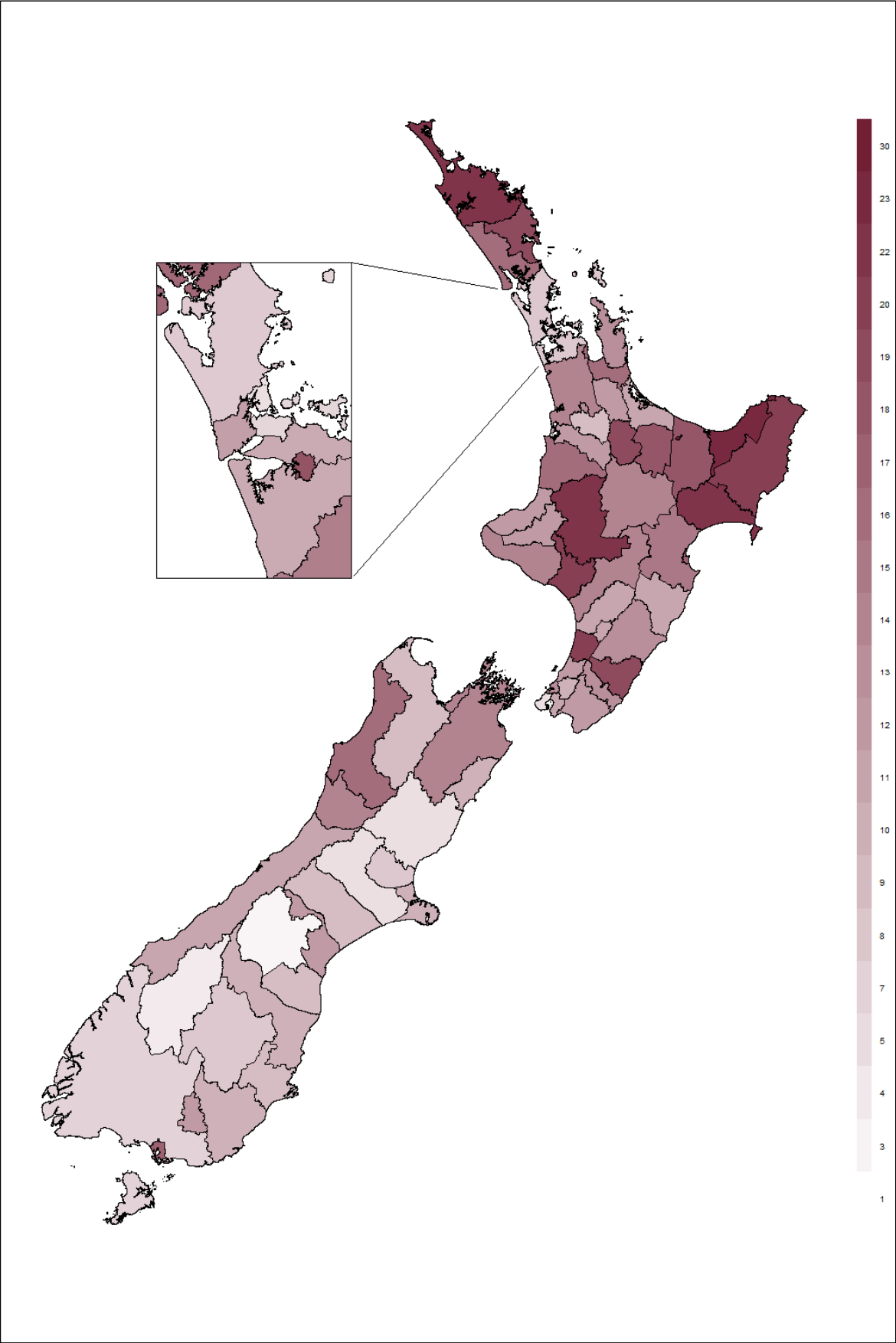
Table 14: Demographic characteristics of the December 2013 youth population by whether at extreme risk of one or more outcomes

	Total 2013 youth population	At extreme risk of one or more poor outcomes	Not at extreme risk of one or more poor outcomes
Number	581,740	61,563	520,173
% Male	51%	54%	51%
% Māori	19%	52%	15%
% European/Pākehā	61%	39%	64%
% Pasifika	9%	8%	9%
% Low SES (NZDep 9 and 10)	24%	45%	21%

Table 15: Estimated outcomes for the 1990/91 birth cohort population by whether at extreme risk of one or more outcomes

	1990/91 birth cohort population	At extreme risk of one or more poor outcomes	Not at extreme risk of one or more poor outcomes
Benefit 5+ years	9.0%	36.0%	5.6%
Corrections sentence	8.7%	33.9%	5.5%
No level 2 by 23	24.7%	62.5%	19.8%
No level 4 by 23	60.4%	88.3%	56.8%
Mental health services	19.7%	42.3%	16.8%

Figure 2: Percentage of the December 2013 youth population at extreme risk of one or more poor outcomes by territorial authority area (Auckland territorial authorities expanded)



The distribution of the at-risk population across New Zealand territorial authority areas is given in Figure 2 (see Appendix 3 Table 1 for the numbers behind the figure). The regions with the largest proportion of their youth population at extreme risk of one or more outcomes are the same as those using the multiple risk measure (Kawerau, Opotiki, Far North and Wairoa districts with 30%, 22%, 22% and 21% predicted to be at risk respectively). As for the alternative risk measure, the largest absolute concentration of at-risk young people is in Auckland, with around one-sixth living in Manukau City, Waitakere City or Papakura District.

Table 16 shows estimated future costs for the at-risk and not-at-risk 1990/91 birth cohort populations at age 20. As with the earlier table showing level of risk across multiple outcomes, there is not a perfect link between estimated risk and expected costs. While most of the not-at-risk population are estimated to have low future costs, almost half of those with high (\$100,000 and over) projected costs are predicted to be not at risk at age 20.

Table 16: Projected welfare and corrections costs for ages 25 to 34 by whether at extreme risk of one or more outcomes at age 20, 1990/91 birth cohort

Projected costs between ages 25 and 34	1990/91 birth cohort population at age 20	At extreme risk of one or more poor outcomes	Not at extreme risk of one or more poor outcomes
\$0	35,004	1,779	33,225
\$1 to \$100,000	20,901	3,348	17,553
\$100,001 to \$200,000	4,200	1,929	2,271
\$200,001+	2,379	1,437	942
Total	62,481	8,493	53,991

Figures 1 to 4 in Appendix 4 show the distribution of costs in greater detail, plotted against the multiple outcome risk measure for the 1990/91 birth cohort population at age 20. Figures 1 and 2 in Appendix 4 show the distribution in terms of absolute numbers of young people (for those with projected costs of less than and more than \$100,000 in total costs between ages 25 and 34 respectively), while Figures 3 and 4 show the distribution expressed as a percentage of young people with each estimated level of risk. These graphs tell a similar story to Table 16 above:

- ▶ In absolute terms, most people with low projected costs have low to moderate estimated risk. At the same time, there are more people with low to moderate estimated risk with projected costs of more than \$100,000 than people with high risk or extreme risk.
- ▶ In relative terms, there is a far higher proportion of people with high or extreme estimated risk with projected costs of greater than \$100,000 or greater than \$200,000, especially in the case of those with estimated extreme risk.

Comparing risk measures

As might be expected given the differences in characteristics of the populations defined by the different risk measures outlined above, the different measures cover slightly different populations. The degree of overlap in the risk populations at age 15 and 21 are examined in Tables 17 and 18 below. In both cases, there is a high degree of overlap between the

populations, with 83% and 82% respectively of the birth cohort population being considered as not being at high risk of poor outcomes using either measure.

Table 17: Estimated multiple outcome risk and extreme risk of one or more poor outcomes at age 15, 1990/91 birth cohort

At extreme risk of one or more poor outcomes	Multiple poor outcome risk level			
	Low to moderate	High	Extreme	Total
No	83.2%	5.0%	0.0%	88.2%
Yes	2.8%	5.0%	4.0%	11.8%
Total	86.0%	10.0%	4.0%	100.0%

Table 18: Estimated multiple outcome risk and extreme risk of one or more poor outcomes at age 21, 1990/91 birth cohort

At extreme risk of one or more poor outcomes	Multiple poor outcome risk level			
	Low to moderate	High	Extreme	Total
No	82.4%	5.7%	0.6%	88.6%
Yes	3.6%	4.3%	3.4%	11.4%
Total	86.0%	10.0%	4.0%	100.0%

An additional 9% and 8% of the 15-year-old and 21-year-old populations respectively were considered to be at risk on both measures. The remaining 8% and 10% of the population respectively were considered to be at risk under one of the outcome measures but not the other. Almost all of those considered to be at extreme risk overall were at extreme risk of at least one poor outcome, particularly at age 15.

How does estimated risk change over time?

We might expect that someone who is predicted to be at risk of poor outcomes at age 15 is likely to be still at risk at older ages, and to some degree, this is true. More than 90% of those who were estimated to have low to moderate risk of poor multiple outcomes at age 15 were still expected to have low to moderate risk at age 21.

As illustrated in Table 19, however, there is some shifting of risk categories over time. Over half of those who were estimated to be at high risk of poor outcomes at age 15 were estimated to have low to moderate risk at age 21, and only one-quarter of those estimated to be at extreme overall risk at age 15 were still estimated as being at extreme risk at age 21. Four-fifths of this latter group were considered to have low to moderate risk at age 21.

Table 19: Estimated multiple outcome risk and extreme risk of one or more poor outcomes at age 15, 1990/91 birth cohort

Multiple poor outcome risk level at age 15	Multiple poor outcome risk level at age 21			
	Low to moderate	High	Extreme	Total
Low to moderate	91.5%	6.6%	1.9%	100.0%
High	57.0%	29.5%	13.5%	100.0%
Extreme	40.4%	35.0%	24.8%	100.0%
Total	86.0%	10.0%	4.0%	100.0%

Table 20 shows a similar story using the risk measure based on extreme risk of one or more poor outcomes. In this case, 94% of those who were not at risk at age 15 were still considered to be not at risk at age 21. On the other hand, only half of those who were estimated to be at risk at age 15 were still considered at risk at age 21. These changes over time are likely to reflect the influence of a number of factors, including new information about a young person’s life becoming available as they get older, changes in people’s lives and circumstances and the challenges inherent in using data of any sort to predict future outcomes.

Table 20: Estimated multiple outcome risk and extreme risk of one or more outcomes at age 15, 1990/91 birth cohort

At extreme risk of one or more poor outcomes at age 15	At extreme risk one or more poor outcomes at age 21		
	No	Yes	Total
No	93.9%	6.1%	100.0%
Yes	49.3%	50.7%	100.0%
Total	88.6%	11.4%	100.0%

5 Target populations

5.1 Approach

For each age group, a cluster analysis was undertaken identifying groups of individuals within the at-risk population, defined as being at extreme risk (top 5% of population risk) of at least one outcome measure. Multiple correspondence analysis was used to redefine the key categorical predictors from the regression modelling into a smaller number of continuous variables, and these were then used to identify a number of clusters at each year of age for females and males jointly.

The youth population was next split into the late teen population (aged 15 to 19) and the early 20s population (aged 20 to 24), and we sought to identify a small number of target populations within each of these age groups. The aim was to identify clearly defined groups at risk of poor outcomes that aligned as closely as possible with the estimated risk from the regression analysis. Target population groupings were informed by the factors that were most predictive of poor outcomes in the regression analysis outlined in the previous section, as well as the clusters identified through the correspondence and cluster analysis, and constructed using the following guiding criteria:

- ▶ Parsimony – target populations should be able to be identified using only a few criteria.
- ▶ Separation – overlap between target populations should be minimised.
- ▶ High sensitivity – most people identified as being at risk should fall into at least one target population.
- ▶ High specificity – most people identified as not being at risk should fall outside of the target populations.

In the end, five groups were identified in each age range. Between them, these groups covered a majority of the at-risk population, and there were no additional clearly identifiable groups that met the criteria above. For the early 20s population, risk was mainly defined using the welfare and corrections outcomes measures, as health and education outcomes could have already occurred at these ages, conflating the risk and outcomes measures.

5.2 Target population descriptions and criteria

The 10 target populations identified are described in Table 21 below along with the criteria by which they can be identified. The two measures of overall risk defined in the previous section are also given. As discussed above, the clusters were primarily designed to align with our at-risk measure, based on someone having extreme risk (being in the top 5% of the population) on at least one outcome measure. Nevertheless, these groups also tend to have a high probability of having high risk across multiple outcomes (our other overall risk categorisation). Around two-thirds or more of each of the 10 target populations are considered to be at extreme risk of one or more poor outcomes, while for most groups, three-

quarters or more are predicted to be at high risk across multiple outcomes. The two groups that are the exception to this are 'Teenagers with health, disability issues or special needs' (aged 15 to 19) and 'Long-term disability beneficiaries' (aged 20 to 24). Both of these groups tend to have extreme risk of a poor welfare and education outcomes, but are far less at risk of poor corrections outcomes than other groups. A breakdown of target populations by territorial authority is included in Appendix 5.

5.3 Target population overlap and coverage

Despite the attempt to identify target populations that not only predict risk well and are identifiable through a few simple criteria but also are mutually exclusive from each other, in practice, these objectives tend to counteract each other, and a trade-off is necessary to reach a balance across the objectives. Figures 3 and 4 below illustrate the degree of overlap between each target population and the other target populations identified in the same age range using Venn diagrams as well as the overlap between the target populations and the at-risk population (those with extreme risk of one or more poor outcomes). These figures relate to the December 2013 population, while overlaps and coverage of the 1990/91 birth cohort population are illustrated in Appendix 6.

The Venn diagrams were designed to be area proportionate such that the area covered by each part of the diagram relates to the size of the populations that meet the relevant criteria. This was accomplished using ellipses instead of circles to represent the different populations.¹⁹ Whilst each Venn diagram is designed to be area proportionate within itself, the diagrams are only broadly comparable with each other, and caution should be taken inferring areas as being equivalent across different diagrams.

Figure 3 shows the various 15 to 19-year-old target populations. Between them, the target populations cover 72% of the total at-risk population, while around 36% of people classified as being in at least one target population do not meet the definition of being at risk. Despite this latter percentage being higher than we might like, these young people may have higher risk than the average person in the not-at-risk population. Between 43% and 60% of the target populations also fall in at least one other target population. The former, with the lowest overlap, is the 'Teenagers with health, disability issues or special needs' group, while the latter, with the highest overlap, is the 'Teenage girls supported by benefits' group.

Figure 4 shows the various 20 to 24-year-old target populations. These target populations cover 82% of the total at-risk population, while 25% of people classified as being in at least one target population do not meet the definition of being at risk. Both of these are an improvement on the younger target populations, reflecting the improving ease of prediction with age. Overlap between target populations is also considerably lower at the older ages, ranging from 7% to 41% (for the 'Long-term disability beneficiaries' and 'Jobseekers in poor health with CYF history' groups respectively).

¹⁹ A software package called eulerAPE was used to represent the areas of overlap in Venn diagram form. See: Luana Micallef and Peter Rodgers (2014). eulerAPE: Drawing Area-proportional 3-Venn Diagrams Using Ellipses. <http://www.eulerdiagrams.org/eulerAPE>

Table 21: Target population descriptions, criteria, size and estimated risk

Target population descriptor		Criteria	Number (2013 popn)	Extreme risk of one or more poor outcomes	High risk across multiple poor outcomes
Age 15 to 19	Teenage boys with Youth Justice or Corrections history	<ul style="list-style-type: none"> Boys aged 18-19 with Corrections history OR Boys aged 15-17 with Youth Justice contact OR Boys aged 15-17 caregiver with custodial history 	12,801	68%	76%
	Teenagers with health, disability issues or special needs	<ul style="list-style-type: none"> Aged 17-19 and on Supported Living Payment Benefit OR Aged 15-19 and used special education services OR Aged 15-19 and attended a special school 	5,769	87%	46%
	Teenage girls supported by benefits	<ul style="list-style-type: none"> Girls aged 15-19 with no qualifications and significant duration on benefit as adult OR Young mothers aged 15-19 on Sole Parent Support Benefit 	4,212	74%	79%
	Mental health service users with stand-down or CYF history	Aged 15-17, used mental health services AND: <ul style="list-style-type: none"> Contact with Child, Youth & Family (CYF) Care and Protection OR History of stand-downs from school 	10,926	82%	81%
	Experienced significant childhood disadvantage	Aged 15-19 AND: <ul style="list-style-type: none"> History of placement in care by Child, Youth & Family OR Notified to CYF with a caregiver with a Corrections history AND supported by benefit for more than 75% of childhood 	16,128	71%	83%
	Not in a target population	None of the above criteria	253,020	5%	7%
Age 20 to 24	Young offenders with custodial sentence	20-24 year olds with a custodial sentence	8,208	86%	88%
	Young offenders with community sentence and CYF history	20-24 year olds with a community sentence (but no custodial sentence) AND: <ul style="list-style-type: none"> A Youth Justice referral OR Notified to CYF 	9,543	72%	78%
	Jobseekers in poor health with CYF history	Received Jobseeker Health Condition, Injury or Disability Benefit for 95% of last year AND: <ul style="list-style-type: none"> Received a Corrections sentence OR referred to Youth Justice OR referred to Child Youth & Family 	2,316	77%	98%
	Sole parents not in fulltime employment with CYF history	Received Sole Parent Support benefit for > 95% of last year AND: <ul style="list-style-type: none"> Received a Corrections sentence OR referred to Youth Justice OR referred to Child, Youth & Family 	6,117	72%	96%
	Long-term disability beneficiaries	20-24 year olds who recieved supported living payment for > 85% of last year	4,521	94%	36%
	Not in a target population	None of the above criteria	264,111	2%	8%

Figure 3: Target population overlaps December 2013 population, ages 15 to 19

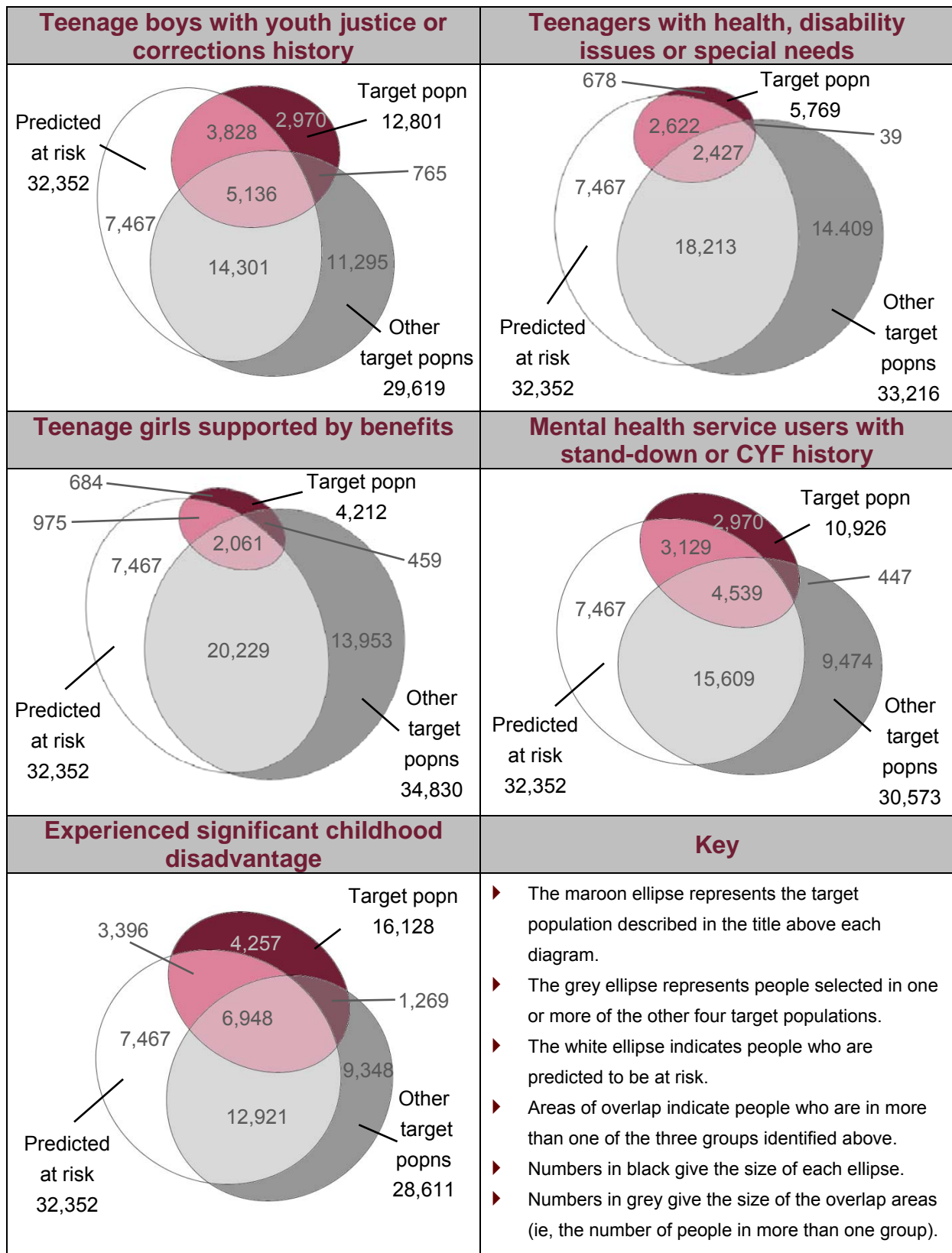
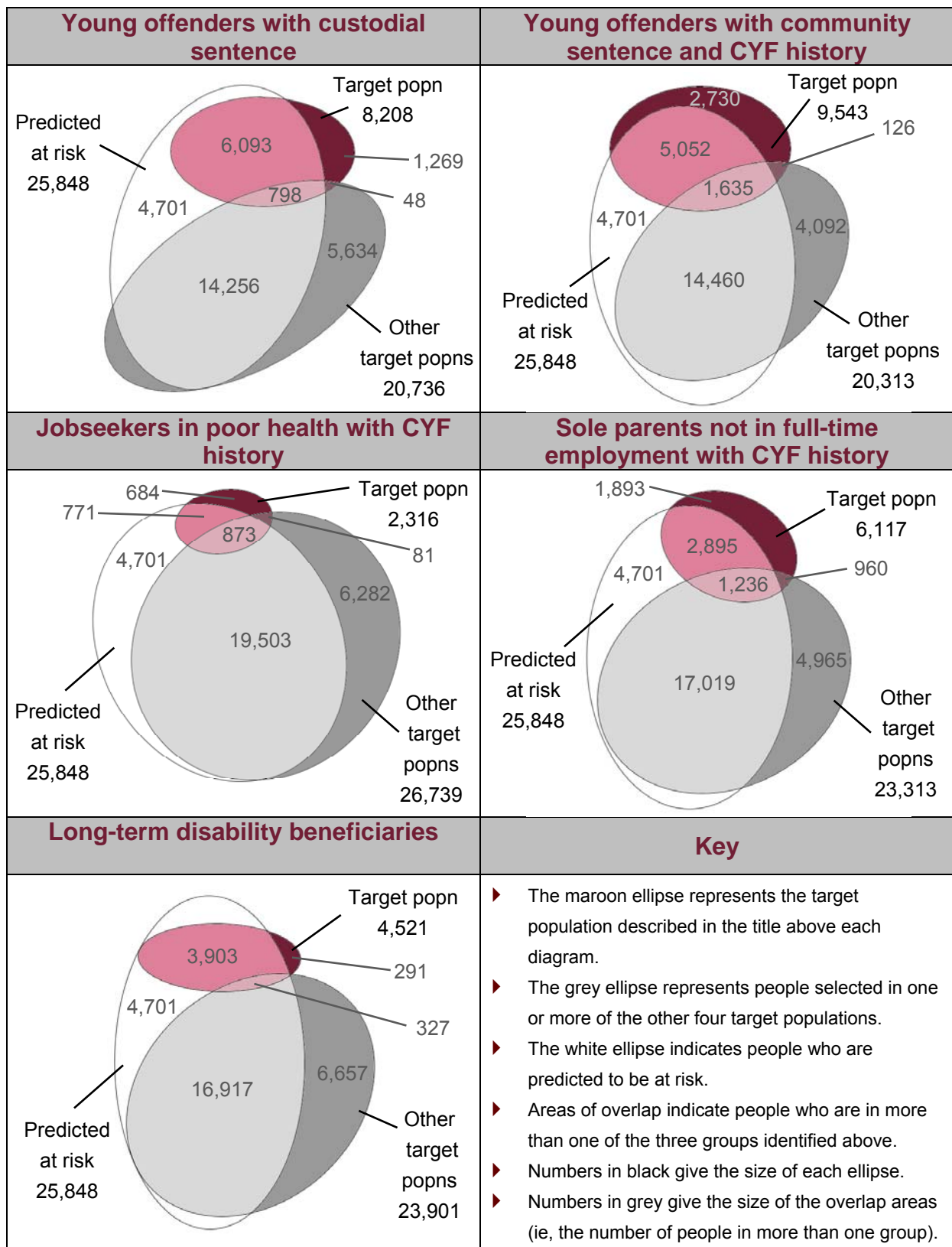


Figure 4: Target population overlaps December 2013 population, ages 20 to 24



5.4 Target population projected outcomes and costs

Expected outcomes for each target population group are outlined in Table 22. As we might expect, young people in target populations are considerably more likely to experience poor outcomes than young people who are not in a target population.

For 15 to 19-year-olds, ‘Teenagers with health, disability issues or special needs’ are most likely to not achieve a level 2 qualification (75%), ‘Mental health service users with stand-down or CYF history’ are most likely to use mental health services (52%), ‘Teenage boys with Youth Justice or Corrections history’ are most likely to receive a corrections sentence (46%), and ‘Teenage girls supported by benefits’ are most likely to be on benefit longterm (62%).

Among 20 to 24-year-olds, ‘Jobseekers in poor health with CYF history’ and ‘Long-term disability beneficiaries’ are most likely to not achieve a level 2 qualification (71%), ‘Jobseekers in poor health with CYF history’ are most likely to use mental health services (75%), ‘Young offenders with a custodial sentence’ are most likely to receive a corrections sentence (67%), and ‘Long-term disability beneficiaries’ are most likely to be on benefit longterm (72%).

Table 22: Expected outcomes by target population

	Target population descriptor	No level 2 quals by age 23	Used mental health service ages 20 to 22	Corrections sentence ages 25 to 34	Longterm benefit receipt ages 25 to 34
Age 15 to 19	Teenage boys with Youth Justice or Corrections history	59%	38%	46%	16%
	Teenagers with health, disability issues or special needs	75%	35%	8%	62%
	Teenage girls supported by benefits	66%	33%	19%	48%
	Mental health service users with stand-down or CYF history	56%	52%	26%	29%
	Experienced significant childhood disadvantage	58%	37%	33%	33%
	Not in a target population	20%	17%	6%	6%
Age 20 to 24	Young offenders with a custodial sentence	63%	61%	67%	29%
	Young offenders with a community sentence and CYF history	62%	44%	52%	29%
	Jobseekers in poor health with CYF history	71%	75%	39%	54%
	Sole parents not in fulltime employment with CYF history	61%	31%	27%	56%
	Long-term disability beneficiaries	71%	43%	9%	72%
	Not in a target population	21%	17%	5%	6%

Average annual expected corrections and benefit costs between ages 25 and 34 are given in Table 23 below for each target population, as well as for the population not covered by a target population at those ages.

Results largely match what we might expect given the population definitions and earlier results. Target populations have higher expected costs than young people not in a target population across all groups. Some groups have particularly high expected corrections costs ('Teenage boys with youth justice or corrections history' and 'Young offenders with custodial sentence') or high income support costs ('Teenagers with health, disability issues or special needs', 'Teenage girls supported by benefits', 'Jobseekers in poor health with CYF history', 'Sole parents not in full-time employment with CYF history' and 'Long-term disability beneficiaries').

While the 'Young offenders with community sentence and CYF history' group is defined by contact with the corrections system, expected corrections costs are not especially high (around \$25,000 per person per annum), and total expected costs are lower than for other target populations at ages 25 to 34. Similarly, at the earlier ages, the 'Teenage boys with youth justice or corrections history' group and the 'Mental health service users with stand-down or CYF history' group have lower expected total costs than other target populations.

In considering this information it is important to recognise that the welfare and corrections costs identified only represent a partial picture of the direct fiscal costs and wider societal costs of poor outcomes.

Table 23: Projected costs by target population group – total from age 25 to 34

	Target population descriptor	Corrections costs (\$000)	Benefit costs age (\$000)	Total projected costs (\$000)
Age 15 to 19	Teenage boys with Youth Justice or Corrections history	50.4	35.4	85.7
	Teenagers with health, disability issues or special needs	7.9	118.1	126.0
	Teenage girls supported by benefits	5.1	110.4	115.5
	Mental health service users with stand-down or CYF history	23.3	62.7	86.0
	Experienced significant childhood disadvantage	30.0	74.9	104.9
	Not in a target population	1.4	16.4	17.8
Age 20 to 24	Young offenders with custodial sentence	101.4	59.8	161.2
	Young offenders with community sentence and CYF history	25.3	66.5	91.8
	Jobseekers in poor health with CYF history	35.6	114.8	150.3
	Sole parents not in fulltime employment with CYF history	6.7	132.5	139.2
	Long-term disability beneficiaries	4.0	132.2	136.2
	Not in a target population	2.0	17.6	19.6

6 Interpreting the A3 document

The accompanying A3 document titled 'Youth at risk: Identifying a target population' presents the results of the analysis in some detail. This section provides some general guidance to interpreting the results.

6.1 Identifying poor long-term outcomes (page 1)

The four outcome measures used to define poor long-term outcomes are outlined under the broad headings of 'Economic opportunity', 'Safety and security', 'Education' and 'Good health'.

6.2 The risk factors most associated with those outcomes (15-year-olds) (page 1)

The regression modelling undertaken in Step 1 described above relates a large number of risk factors to each of the four outcome measures used. The statistical strength of this relationship can be assessed according to the order in which the factors were selected by the forward selection procedure in the regression modelling – this procedure progressively adds factors to the model according to the additional explanatory power that factor contributes to the model at each stage of selection.²⁰

The factors listed in the A3 document present an example of those five factors that are considered to be most predictive of each outcome using age 15 as an example. These are the factors added earliest in the modelling procedure for the regression models of 15-year-olds (and therefore that explain the most variation in outcomes, conditional on other variables already added). These are compared across models for both females and males, with extra weighting afforded factors that are highly predictive for both females and males. In saying this, all factors listed were significant predictors for both males and females. This list should be considered broadly indicative of the factors that are most important in predicting poor future outcomes for 15-year-olds.

Factors are different for different outcomes, but some factors are highly predictive across multiple outcomes. Being notified to CYF as a child was highly predictive of poor outcomes across all four domains, while ethnicity was significant across three domains (all except for having no level 2 qualifications by age 23), as was being stood down from school (all except for being on a benefit for more than five years). Having a caregiver with benefit receipt and/or low qualifications, receiving special education services and spending a long time on a benefit as a child were all highly predictive across two outcomes areas.

²⁰ At each stage of the procedure, the process examines the score chi-squared statistic for each factor were it to be added individually to the existing model. The factor with the highest chi-squared score is added and the procedure repeated until there are no remaining factors with chi-squared scores that are statistically significant at the 5% level of significance.

These findings should be interpreted with some caution for a number of reasons. Most importantly, whilst the association between a factor and a future outcome means that that factor may be a useful predictor of future outcomes, it does not necessarily mean there is a causal relationship between the two.

Additionally, factors are identified as being highly predictive in the modelling if they add something on top of the factors already selected for the model. Where a number of factors are highly correlated with each other, only one may be selected for the model even though the relationships may be complex, and correlated factors may also be independently highly predictive and bear an important relationship to the outcome of interest.

Finally, we have a limited set of observed predictive factors we can use from administrative data. In many cases, these factors may merely be acting as a proxy for other, unobserved factors that we are unable to measure. As young people enter their adult years, more information becomes available that can be used to determine their risk of poor outcomes. In many cases, this is a direct early indicator of the outcome of interest (for example, long-term benefit receipt in the late teen years is a direct measure of early long-term benefit receipt – the ‘economic opportunity’ outcome measure).

6.3 Identifying those most at risk (page 1)

As discussed above, the regression modelling allows estimated risk scores of future poor outcomes to be calculated for a ‘current’ population’ as at 31 December 2013. An estimated risk score of poor outcomes across multiple domains was calculated using a person’s average ranked risk across the four domains. These average ranks were themselves ranked and the top 5% of young people selected and categorised as being at extreme risk, with the next 10% of individuals categorised as high risk.

The table in the A3 document contrasts the demographic characteristics and projected future outcomes of these high-risk groups with the rest of the youth population and the total youth population. Extreme-risk individuals were more likely to:

- ▶ receive a benefit for more than five years between ages 25 and 34 – 40% compared to 26% of those at high risk and 5% of youth not identified as high risk
- ▶ be given a custodial or community sentence between ages 25 and 34 – 46% compared to 24% of those identified as high risk and 5% of those not at high risk
- ▶ not achieve a level 2 qualification by age 23 (72% compared to 58% of those at high risk and 18% of other young people) or a level 4 qualification by age 23 (95% compared to 90% of those at high risk and 55% of other young people)
- ▶ use mental health or addiction services or mental health pharmaceuticals when aged 20 to 22 – 58% compared to 36% of those at high risk and 16% of those not at high risk.

High-risk individuals were more likely to be Māori and to live in areas of relatively high deprivation and were likely to have higher future corrections and benefit costs. The map and table at the bottom of the page shows that, whilst over a quarter of high-risk youth live in Auckland, young people outside of the main centres tend to be more likely to be high risk, particularly those living in the Gisborne, Northland, Hawke’s Bay or Bay of Plenty regions.

6.4 Characteristics of at-risk groups by age (page 2)

The demographic characteristics and risk factors associated with poor outcomes change with age, as illustrated on page 2 of the A3 document. Predictors that were important at age 15 are not necessarily as important at age 20, as other potentially more predictive factors become available. Whilst indicators such as being stood down from school, having a CYF notification as a child or having been supported by a benefit for a long time as a child were predictive of poor outcomes at age 15, by age 20, indicators related to personal experience in the corrections system (through being sentenced), long-term benefit receipt or time out of employment, education or training (ie, NEET), using mental health services and a lack of qualifications to that point became most important. These latter measures are more closely aligned with the outcomes measures used and could be seen as direct early indicators of these outcomes.

6.5 Target populations 15 to 19 years and 20 to 24 years (pages 3 to 6)

Targeting investment toward those who need it most based on an individualised risk measure is often difficult to accomplish, either due to practical considerations, such as the efficiency of focusing efforts at a distinct geographic area or community, or due to data limitations that could restrict the ability to calculate individualised risk. As such, it may be necessary for investment to be targeted at specific target populations identified through a smaller set of identifiable characteristics. Some potential target populations (five for people aged 15 to 19 and five for people aged 20 to 24) were identified based on regression modelling, clustering and descriptive analysis using the approach outlined above.

These populations are described in pages 3 to 6 of the A3 document – with pages 3 and 5 describing the demographic characteristics and outcomes for each population at ages 15 to 19 and 20 to 24 respectively and pages 4 and 6 indicating their geographic location.

This represents one potential way the model could be used to target services. Numerous other approaches will be equally valid, depending on the nature of investment being considered. Although an attempt was made to make the target populations as separate as possible while still being highly predictive of risk, there is some overlap between target populations, especially in the 15 to 19 age range, where the construction of the target populations relies on a wide range of characteristics.

In broad terms, the five target populations at ages 15 to 19 are constructed around a mixture of age and gender criteria alongside childhood indicators of risk such as contact with CYF or youth justice and, in the late teens, benefit receipt. Target populations in the 20 to 24 age range are largely specified on a smaller number of criteria defined largely by a history of benefit receipt and corrections sentences. This is both consistent with the fact that only corrections and welfare outcomes are able to be modelled for most of these ages and that benefit type is related to other domains, such as ill health and disability.

7 Conclusions

This paper has presented findings from an analysis of Statistics New Zealand's Integrated Data Infrastructure. The analysis looked at the characteristics of young people aged 15 to 24 who were at risk of poor outcomes as adults across welfare, corrections, education and health domains, and attempted to define useful and identifiable target populations at high risk of experiencing these poor outcomes.

The work was undertaken by the Analytics and Insights team at the Treasury in collaboration with other government agencies. It fed into a stream of work being led by the Ministry of Education, resulting in the production of the accompanying A3 document. The analysis presented here represents one of a number of early steps towards using a more data-driven approach to prioritising social assistance initiatives and evaluating the effectiveness of social assistance programmes. The analysis has highlighted a number of characteristics that are predictive of future poor outcomes. Examples include early contact with government agencies such as Child, Youth and Family (CYF), caregiver qualifications and benefit status, geographic location and participation and early outcomes in the education system. These can be used to quantify risk at an individual level and to identify the size and characteristics of at-risk groups of young people at different ages.

The characteristics that are predictive of future outcomes change over time. As young people progress into early adulthood, poor future outcomes become directly evident through contact with the benefit, corrections and health systems. This, combined with the proximity of the outcomes period we are seeking to predict, means that it becomes easier to predict poor outcomes as a young person ages. At the same time, however, these outcomes may become more and more difficult to influence.

It is possible to identify groups of at-risk youth at different ages using a small set of identifying characteristics, such as benefit type and duration, corrections sentencing information and information on a person's early contact with government agencies such as CYF. These predictions are by no means perfect however. Those young people who are identified as being at risk are much more likely to have poor future outcomes than those who aren't, but a large number of people have poor outcomes despite not falling into one of these defined groupings. Approaches to targeting services should be flexible enough to offer services based on particular individual circumstances as well as broad characteristics.

One useful way of targeting services is to focus on specific areas with higher concentrations of at-risk youth. However, there is a tension between targeting those services at areas where a high proportion of youth are at risk (such as Kawerau or Opotiki) and larger centres where large number of at-risk youth live (such as Manukau or Waitakere).

All of the findings in the paper should be treated with some caution given the various caveats associated with the data and methods used as well as the early stage of this type of analysis. There is some scope to improve the results in future, taking advantage of the improvements that are being made to linked administrative data and refining the analytical and estimation methods.

Appendices

Appendix 1: Predictive factors at different ages

The table below outlines the number of models in which a predictor was selected for inclusion in the modelling at each age.²¹ Each predictor could be used in up to eight models (up to age 20), one for each of the four outcomes for both males and females. A zero in the table indicates that the factor was not included in any of the regression models at that age, while a missing value indicates that the factor was not available for inclusion, generally because that type of information was not available at that age.

Appendix 1 Table 1: Predictive factors and number of models by age

Risk factor	Age							
	15	16	17	18	19	20	21*	22*†
Socio-demographic and location characteristics								
Ethnicity	8	8	8	8	8	8	6	4
Territorial authority	6	5	5	4	4	4	1	0
New Zealand Deprivation Index (NZDep) decile	7	7	6	5	3	3	1	0
Currently overseas	0	1	0	1	3	3	1	2
Childhood risk factors								
Notified to CYF care and protection as a child	8	8	8	8	8	8	5	3
CYF care and protection maltreatment finding	2	2	1	1	2	1	2	1
Placed in care of CYF care and protection	6	6	4	3	0	2	0	0
Referred to youth justice	7	8	7	6	6	6	4	4
Placed in care by youth justice	2	2	2	2	1	1	1	1
Maternal caregiver education/benefit status	8	8	8	8	8	6	5	2
Caregiver with community sentence	7	8	7	7	4	5	2	1
Caregiver with custodial history	4	1	0	0	1	1	1	0
Duration on a benefit as a child	5	3	3	2	3	2	0	0
Main type of benefit as a child	3	3	3	3	1	2	0	1
Schooling characteristics								
Enrolled at school	0	0	7	5	4	2	2	0
Type of school (private, state etc.)	0	0	0	0	0	0	0	0
Currently in special school	4	4	1	0	0	1	0	0
Ever received special education services	5	5	5	5	5	4	3	3
Ever stood down from school	8	7	7	7	6	6	3	0
Ever suspended from school	7	7	4	4	2	4	1	1
Ever truant from school	2	6	2	2	1	2	1	0
Last school decile	6	6	6	4	3	3	1	0
Last school was correspondence	1	1	0	0	1	1	0	0
Last school was private	1	2	1	3	0	1	1	0
Achieved level 1 NCEA or equivalent	0	8	3	2	0	0	0	0
Achieved level 2 NCEA or equivalent	0	2	2	5	4	4	1	0
Achieved level 3 NCEA or equivalent	0	0	1	3	3	3	1	1

²¹ As discussed in the body of the paper, only those predictors that were statistically significant based on a forward selection approach were included in each model.

Risk factor	Age							
	15	16	17	18	19	20	21*	22*†
Tertiary education								
Enrolled for 1 year in tertiary education				1	3	1	3	0
Enrolled for 2 years in tertiary education				1	1	3	2	0
Enrolled for 3 years in tertiary education					2	0	3	1
Latest tertiary enrolment level			3	4	4	3	3	0
Highest school or tertiary level completed	1	0	5	4	3	3	2	0
Highest tertiary enrolment level			0	3	6	5	5	2
Employment and welfare characteristics								
Duration not in employment, education or training (NEET) last year			8	2	3	2	2	1
Duration NEET since 16			0	8	8	6	4	1
Duration NEET since 18					0	2	2	1
Ever on a benefit as an adult			3	5	1	4	1	0
Main benefit type since age 18				1	5	4	2	2
Main benefit type last year			5	4	4	3	6	3
Duration on a benefit since age 18				0	2	2	4	4
Duration on a benefit last year			0	0	3	2	4	3
On a benefit for 2 years as adult			0	1	0	2	1	0
Any earnings in last 2 years (wages and salaries and self-employment)	1	1	2	2	0	3	1	0
Any earnings in last year	4	4	1	0	2	0	0	1
Average earnings in last 2 years	1	0	2	3	3	3	2	0
Total earnings in last year	1	5	3	4	5	4	3	1
Corrections contact								
Served community sentence				2	2	1	1	0
Served custodial sentence				0	0	1	0	0
Served some sort of corrections sentence				5	7	6	5	3
Use of mental health services								
Used alcohol or drug addiction services	0	0	0	1	0	1	0	0
Used any mental health services	0	0	0	0	0	0	0	0
Used other mental health services	0	0	0	5	4	4	1	2
Indicator of any mental health illness	1	1	1	1	2	1	1	0
Indicator of other mental health illness	7	7	7	7	5	7	2	0
Indicator of substance abuse	2	3	2	2	2	2	1	1
Early parenting and offspring childhood risk factors								
Parent	0	0	1	1	1	1	1	0
Early parent (before age 19)	0	0	0	0	0	0	0	0
Had own child in placement or with maltreatment finding		0	0	0	0	0	0	0
Had own child in Police/family violence notification		2	1	1	3	2	1	0
Had own child placed in CYF care		1	1	1	1	1	1	0
Had own child with maltreatment finding		2	1	1	1	2	1	1
Own child referred to youth justice								
Number of models run*	8	8	8	8	8	8	6	4
Average factors included per model	15.6	18	18.4	21	20.5	20.5	18	12.8
Potential factors	42	46	54	61	63	63	63	63

* At age 21, only three outcomes were modelled, and factors could be included in only six models (rather than eight). At age 22, only two outcomes were modelled, and factors could be included in only four models.

† The 22-year-old models were also applied to the population at ages 23 to 25.

Appendix 2: Outcomes by territorial authority

Appendix 2 Table 1: Estimated outcomes at ages 25 to 34 for youth 1990/91 cohort by the territorial authority of residence at age 15

Territorial authority			Outcomes				
	Cohort number	Cohort %	No level 2 quals	No level 4 quals	Mental health	Corrections sentence	Long-term benefit
Ashburton District	420	1%	21%	56%	23%	6%	6%
Auckland City	5,349	9%	21%	54%	16%	7%	7%
Banks Peninsula District	84	0%	18%	68%	29%	s	s
Buller District	153	0%	41%	82%	22%	12%	10%
Carterton District	132	0%	25%	64%	34%	9%	11%
Central Hawke's Bay District	219	0%	32%	73%	21%	7%	8%
Central Otago District	219	0%	18%	49%	22%	7%	7%
Christchurch City	4,683	7%	26%	58%	23%	7%	8%
Clutha District	255	0%	25%	62%	20%	9%	6%
Dunedin City	1,563	2%	21%	56%	24%	9%	7%
Far North District	966	2%	30%	72%	19%	15%	15%
Franklin District	930	1%	27%	57%	19%	7%	8%
Gisborne District	792	1%	32%	73%	22%	15%	13%
Gore District	213	0%	21%	65%	23%	13%	8%
Grey District	174	0%	26%	71%	28%	9%	7%
Hamilton City	2,028	3%	25%	57%	23%	8%	11%
Hastings District	1,248	2%	27%	62%	21%	9%	11%
Hauraki District	288	0%	38%	72%	21%	13%	9%
Horowhenua District	483	1%	30%	75%	18%	10%	13%
Hurunui District	165	0%	16%	53%	20%	5%	5%
Invercargill City	822	1%	28%	64%	23%	11%	8%
Kaikoura District	45	0%	47%	73%	33%	20%	s
Kaipara District	300	0%	24%	70%	17%	10%	8%
Kapiti Coast District	717	1%	23%	59%	22%	10%	10%
Kawerau District	141	0%	38%	79%	26%	19%	23%
Lower Hutt City	1,626	3%	23%	62%	18%	9%	10%
Mackenzie District	63	0%	s	38%	19%	s	s
Manawatu District	456	1%	21%	64%	20%	9%	8%
Manukau City	5,709	9%	28%	62%	15%	9%	10%
Marlborough District	612	1%	23%	65%	23%	8%	8%
Masterton District	390	1%	22%	64%	27%	10%	12%
Matamata-Piako District	531	1%	24%	60%	25%	11%	7%
Napier City	936	1%	26%	64%	23%	11%	10%
Nelson City	723	1%	24%	59%	24%	9%	8%
New Plymouth District	1,167	2%	24%	60%	21%	7%	7%
North Shore City	3,141	5%	17%	49%	17%	4%	4%
Opotiki District	198	0%	35%	83%	18%	15%	12%
Otorohanga District	150	0%	20%	58%	16%	14%	12%
Palmerston North City	1,119	2%	23%	64%	21%	9%	9%
Papakura District	819	1%	32%	64%	18%	12%	13%
Porirua City	882	1%	30%	67%	18%	11%	13%
Queenstown-Lakes District	219	0%	16%	53%	18%	5%	7%
Rangitikei District	267	0%	28%	78%	21%	10%	8%
Rodney District	1,299	2%	23%	55%	22%	7%	8%
Rotorua District	1,188	2%	30%	63%	22%	14%	12%
Ruapehu District	216	0%	35%	81%	18%	17%	13%
Selwyn District	465	1%	17%	50%	17%	5%	6%
South Taranaki District	432	1%	34%	74%	21%	10%	12%

Territorial authority	Outcomes						
	Cohort number	Cohort %	No level 2 quals	No level 4 quals	Mental health	Corrections sentence	Long-term benefit
South Waikato District	429	1%	31%	76%	21%	15%	12%
South Wairarapa District	123	0%	24%	68%	24%	7%	10%
Southland District	417	1%	22%	60%	19%	8%	7%
Stratford District	153	0%	25%	55%	16%	4%	10%
Taranua District	327	1%	27%	73%	18%	8%	13%
Tasman District	705	1%	22%	60%	21%	8%	7%
Taupo District	531	1%	32%	66%	19%	10%	7%
Tauranga District	1,581	3%	20%	55%	20%	10%	9%
Thames-Coromandel District	372	1%	27%	61%	23%	9%	8%
Timaru District	720	1%	21%	59%	23%	6%	9%
Upper Hutt City	645	1%	23%	63%	22%	7%	9%
Waikato District	801	1%	28%	67%	18%	10%	12%
Waimakariri District	678	1%	26%	60%	20%	8%	7%
Waimate District	126	0%	19%	64%	19%	10%	7%
Waipa District	729	1%	26%	57%	19%	5%	9%
Wairoa District	150	0%	38%	90%	20%	22%	24%
Waitakere City	3,078	5%	27%	61%	19%	8%	10%
Waitaki District	297	0%	17%	55%	20%	7%	6%
Waitomo District	162	0%	39%	74%	20%	11%	17%
Wanganui District	771	1%	28%	67%	22%	15%	13%
Wellington City	2,085	3%	13%	50%	19%	5%	6%
Western Bay of Plenty District	768	1%	20%	55%	19%	9%	7%
Westland District	126	0%	31%	69%	21%	10%	10%
Whakatane District	663	1%	30%	67%	19%	12%	13%
Whangarei District	1,299	2%	27%	66%	20%	13%	12%

s = suppressed for confidentiality reasons.

Appendix 3: Estimated risk by territorial authority, 2013 youth population

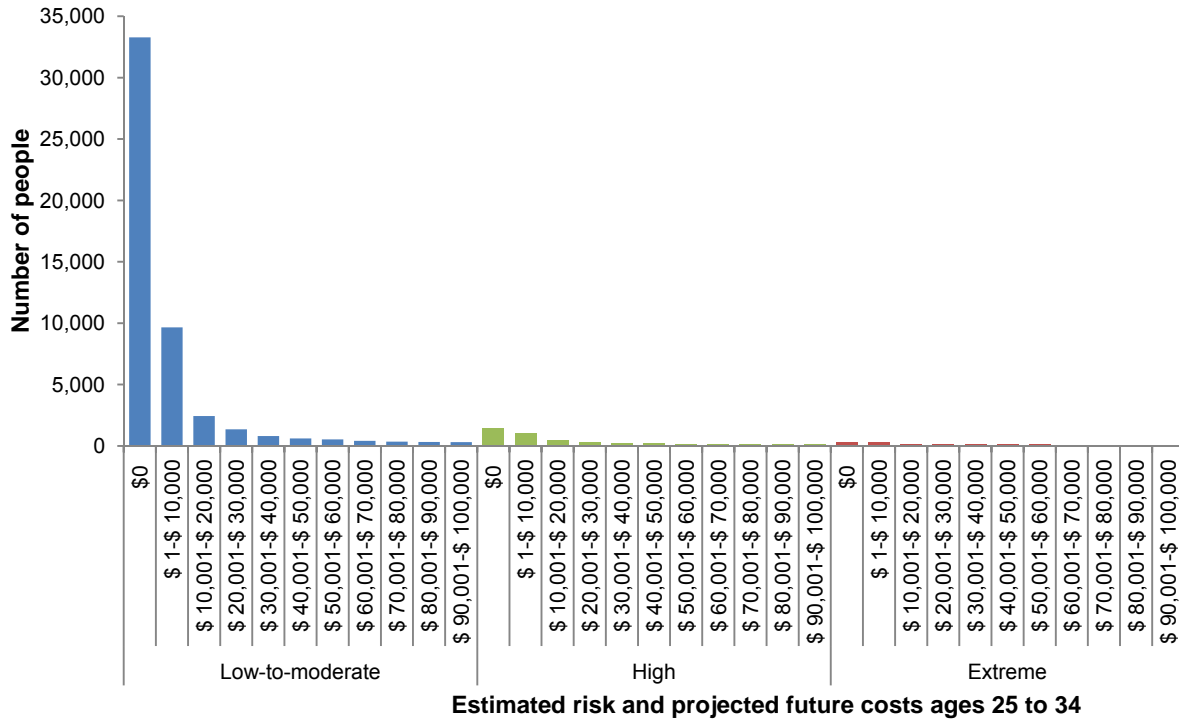
Appendix 3 Table 1: Estimated risk for December 2013 population aged 15 to 24 by territorial authority of residence

Territorial Authority	Total population aged 15 to 24	High risk across multiple outcomes	At extreme risk of one or more poor outcomes
Ashburton District	3,756	11%	9%
Auckland (combined)	200,790	11%	8%
Auckland City	55,947	8%	6%
Franklin District	8,712	14%	9%
Manukau City	57,219	14%	10%
North Shore City	29,964	5%	4%
Papakura District	7,866	26%	18%
Rodney District	12,462	10%	6%
Waitakere City	28,620	15%	10%
Buller District	1,143	20%	16%
Carterton District	843	12%	12%
Central Hawkes Bay District	1,506	13%	10%
Central Otago District	1,641	5%	8%
Christchurch City	47,505	13%	10%
Clutha District	2,103	13%	9%
Dunedin City	18,873	11%	9%
Far North District	7,374	30%	22%
Gisborne District	6,759	29%	20%
Gore District	1,671	15%	12%
Grey District	1,830	20%	14%
Hamilton City	22,224	19%	13%
Hastings District	10,182	22%	15%
Hauraki District	2,247	22%	15%
Horowhenua District	3,789	27%	19%
Hurunui District	1,158	2%	4%
Invercargill City	7,212	23%	16%
Kaikoura District	360	3%	9%
Kaipara District	2,304	21%	15%
Kapiti Coast District	5,493	17%	11%
Kawerau District	978	42%	30%
Lower Hutt City	13,740	17%	11%
Mackenzie District	357	0%	2%
Manawatu District	3,873	15%	11%
Marlborough District	4,794	18%	14%
Masterton District	3,111	25%	19%
Matamata-Piako District	4,314	16%	11%
Napier City	7,707	22%	16%
Nelson City	5,970	18%	14%
New Plymouth District	9,540	16%	11%
Opotiki District	1,287	31%	22%

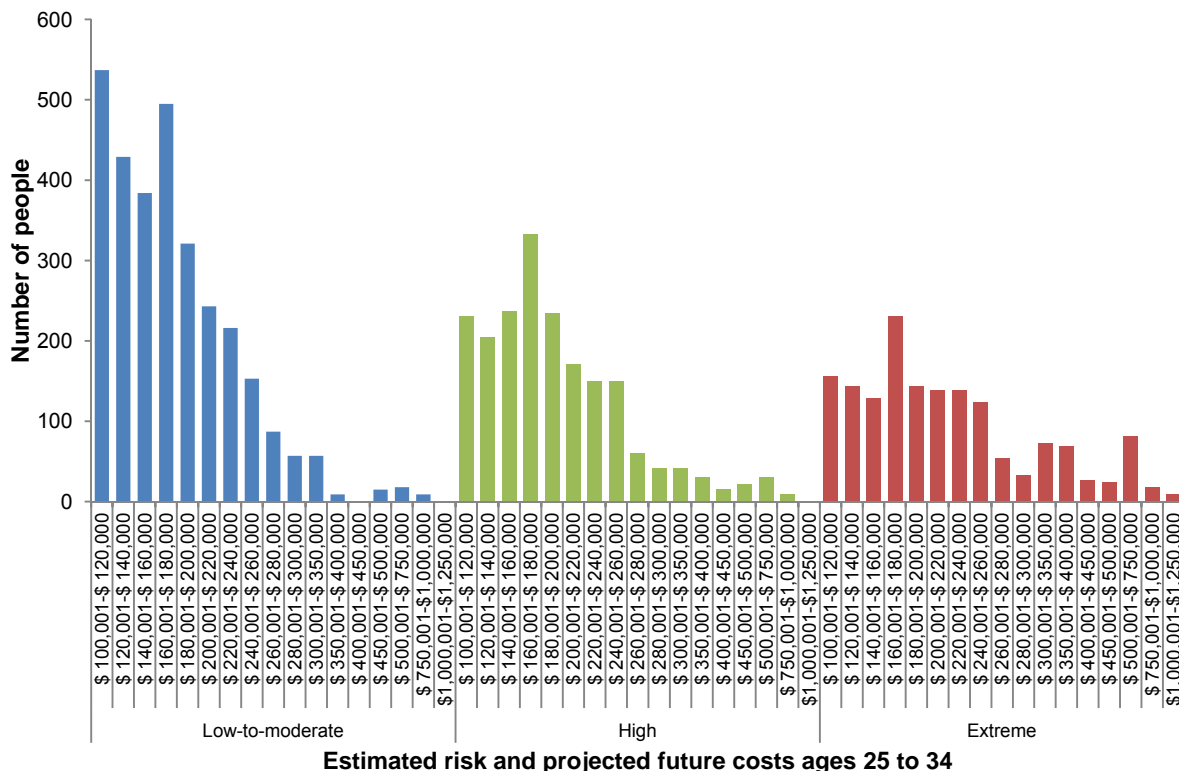
Territorial Authority	Total population aged 15 to 24	High risk across multiple outcomes	At extreme risk of one or more poor outcomes
Otorohanga District	1,203	13%	12%
Palmerston North City	12,165	15%	11%
Porirua City	7,731	19%	13%
Queenstown-Lakes District	2,448	1%	3%
Rangitikei District	1,875	18%	13%
Rotorua District	9,948	25%	18%
Ruapehu District	1,692	26%	21%
Selwyn District	5,046	4%	4%
South Taranaki District	3,693	21%	14%
South Waikato District	3,441	28%	18%
South Wairarapa District	1,005	15%	11%
Southland District	3,549	10%	7%
Stratford District	1,257	13%	12%
Tararua District	2,268	19%	13%
Tasman District	5,391	12%	9%
Taupo District	4,176	20%	14%
Tauranga District	14,178	18%	12%
Thames-Coromandel District	2,499	18%	12%
Timaru District	5,625	15%	11%
Upper Hutt City	5,487	14%	9%
Waikato District	6,576	19%	13%
Waimakariri District	6,180	10%	8%
Waimate District	828	7%	9%
Waipa District	6,384	12%	9%
Wairoa District	1,044	30%	21%
Waitaki District	2,316	14%	9%
Waitomo District	1,248	18%	16%
Wanganui District	5,970	28%	20%
Wellington City	26,919	5%	4%
Western Bay of Plenty District	5,595	15%	11%
Westland District	717	6%	11%
Whakatane District	4,734	27%	17%
Whangarei District	10,629	27%	19%
Total New Zealand	581,740	11%	15%

Appendix 4: Distribution of projected costs by estimated risk of poor outcomes at age 20

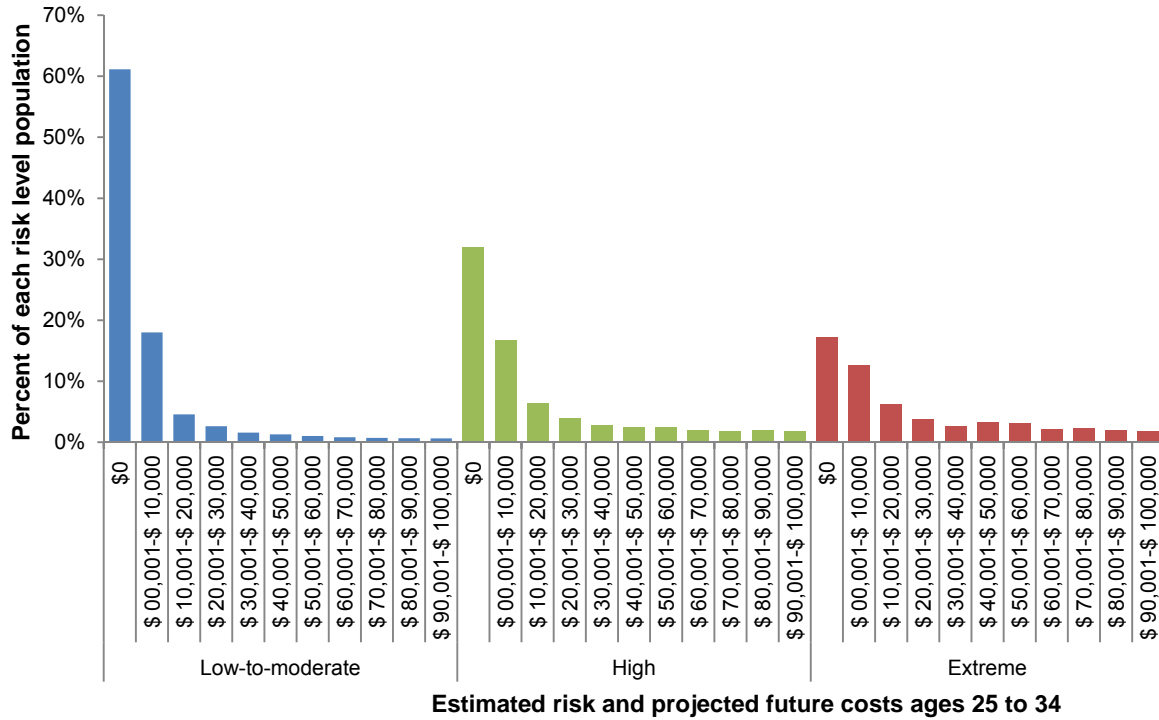
Appendix 4 Figure 1: Projected welfare and corrections cost distribution by risk level at age 20, 1990/91 birth cohort (numbers of people – \$0 to \$100,000 only)



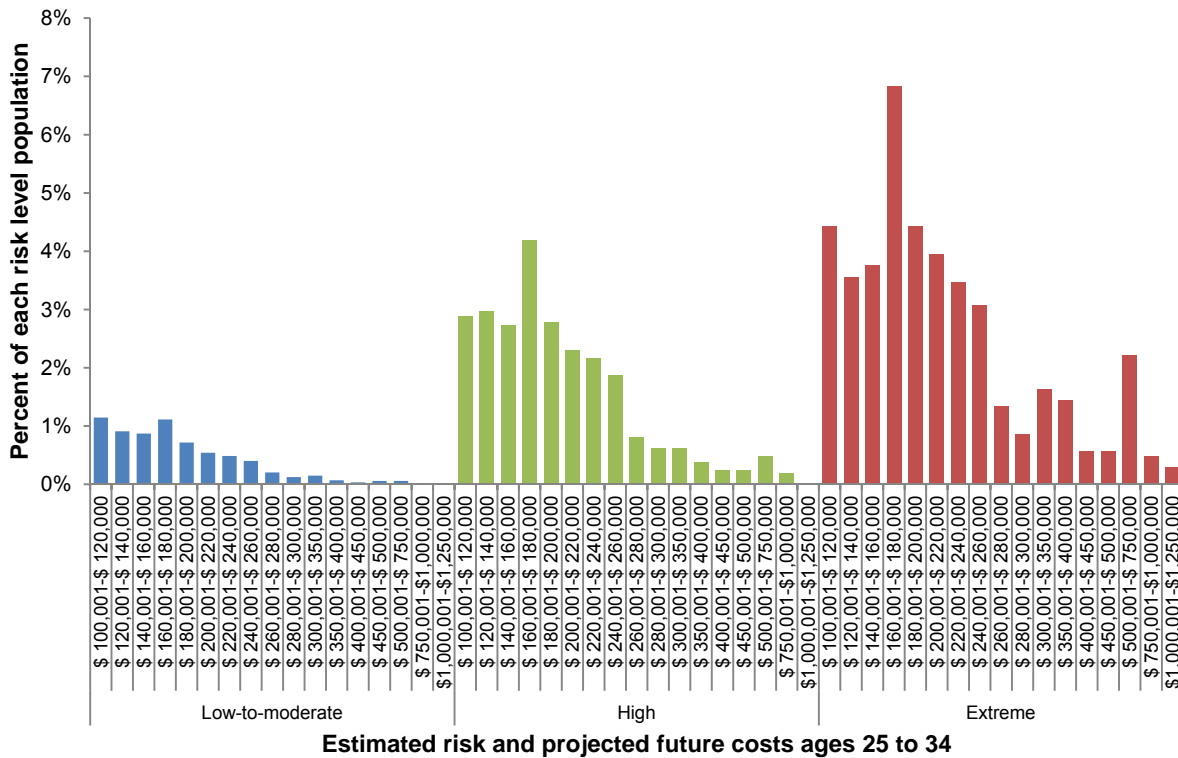
Appendix 4 Figure 2: Projected welfare and corrections cost distribution by risk level at age 20, 1990/91 birth cohort (numbers of people – over \$100,000 only)



Appendix 4 Figure 3: Projected welfare and corrections cost distribution by risk level at age 20, 1990/91 birth cohort (percentage of each risk level population – \$0 to \$100,000 only)



Appendix 4 Figure 4: Projected welfare and corrections cost distribution by risk level at age 20, 1990/91 birth cohort (percentage of each risk level population – over \$100,000 only)



Appendix 5: Target populations by territorial authority, 2013 youth population

Appendix 5 Table 1: Target populations by territorial authority for December 2013 population (ages 15 to 19)

	Teenage boys with Youth Justice or Corrections history	Teenagers with health, disability issues or special needs	Teenage girls supported by benefits	Mental health service users with stand-down or CYF history	Experienced significant childhood disadvantage	In any target population	Not in a target population
Ashburton District	66	24	15	57	72	186	1,782
Auckland (combined)	3,321	1,830	1,320	2,940	4,419	10,350	88,134
Auckland City	675	441	255	618	858	2,151	23,247
Franklin District	201	87	60	153	213	543	4,248
Manukau City	1,200	534	522	972	1,665	3,591	25,134
North Shore City	276	210	69	312	279	912	13,908
Papakura District	261	117	138	213	411	810	3,024
Rodney District	177	117	48	183	186	561	6,243
Waitakere City	531	324	228	489	807	1,782	12,330
Banks Peninsula District	9	s	s	9	15	27	414
Buller District	33	12	6	36	36	96	492
Carterton District	21	9	6	30	27	66	438
Central Hawkes Bay District	48	9	9	24	39	99	747
Central Otago District	36	18	12	33	33	102	852
Christchurch City	861	402	291	792	1,083	2,466	18,495
Clutha District	57	27	12	63	54	159	1,011
Dunedin City	279	147	60	345	342	849	6,699
Far North District	357	72	126	297	492	957	3,237
Gisborne District	264	90	93	234	330	720	2,904
Gore District	51	24	9	27	36	117	783
Grey District	57	27	12	72	54	153	774
Hamilton City	552	240	219	357	816	1,527	8,208
Hastings District	333	90	105	243	408	852	4,797
Hauraki District	72	48	27	69	99	210	1,029
Horowhenua District	135	57	48	129	195	414	1,665

	Teenage boys with Youth Justice or Corrections history	Teenagers with health, disability issues or special needs	Teenage girls supported by benefits	Mental health service users with stand-down or CYF history	Experienced significant childhood disadvantage	In any target population	Not in a target population
Hurunui District	18	9	9	15	24	57	645
Invercargill City	267	126	54	180	225	588	2,904
Kaikoura District	18	9	s	9	21	36	192
Kaipara District	87	21	30	69	114	231	1,098
Kapiti Coast District	96	60	24	129	159	348	2,715
Kawerau District	57	21	21	57	87	162	381
Lower Hutt City	240	123	111	279	405	837	5,910
Mackenzie District	s	s	s	6	s	15	213
Manawatu District	75	51	24	75	111	261	1,887
Marlborough District	150	78	36	129	156	393	2,229
Masterton District	90	57	24	99	141	300	1,350
Matamata-Piako District	111	60	27	72	111	282	1,974
Napier City	237	69	81	192	342	675	3,492
Nelson City	123	69	42	150	216	438	2,739
New Plymouth District	264	120	66	210	297	714	4,359
Opotiki District	60	21	24	45	87	168	519
Otorohanga District	36	12	9	12	33	75	570
Palmerston North City	252	141	87	240	351	753	4,449
Porirua City	174	81	66	141	246	546	3,378
Queenstown-Lakes District	27	9	s	30	12	69	1,179
Rangitikei District	45	24	15	42	45	129	864
Rotorua District	435	141	120	240	462	1,005	4,320
Ruapehu District	78	24	18	30	99	183	732
Selwyn District	51	42	15	63	36	171	2,640
South Taranaki District	111	42	39	84	123	285	1,656
South Waikato District	135	48	54	87	147	357	1,542
South Wairarapa District	27	15	12	30	39	87	525
Southland District	87	42	18	48	60	207	1,710
Stratford District	45	18	9	24	39	99	573
Tararua District	63	30	18	51	105	183	1,053

	Teenage boys with Youth Justice or Corrections history	Teenagers with health, disability issues or special needs	Teenage girls supported by benefits	Mental health service users with stand-down or CYF history	Experienced significant childhood disadvantage	In any target population	Not in a target population
Tasman District	99	81	27	135	123	345	2,829
Taupo District	165	57	36	93	171	387	1,884
Tauranga District	444	120	93	351	435	1,044	6,585
Thames-Coromandel District	69	42	21	66	96	213	1,167
Timaru District	147	45	33	168	144	393	2,628
Upper Hutt City	78	60	42	90	123	309	2,547
Waikato District	159	87	66	111	273	507	3,102
Waimakariri District	114	48	21	93	90	279	3,207
Waimate District	24	12	9	27	30	72	450
Waipa District	111	69	42	99	123	342	3,204
Wairoa District	45	15	15	27	51	114	471
Waitaki District	42	18	15	63	57	141	1,131
Waitomo District	33	12	12	12	45	96	609
Wanganui District	243	69	66	144	333	624	2,610
Wellington City	192	153	60	213	225	636	10,017
Western Bay of Plenty District	156	51	30	132	168	390	2,796
Westland District	24	6	6	21	30	66	366
Whakatane District	213	57	57	183	231	501	2,133
Whangarei District	414	93	129	312	528	1,035	4,647

s = suppressed for confidentiality reasons.

Appendix 5 Table 2: Target populations by territorial authority for December 2013 population (ages 20 to 24)

	Young offenders with custodial sentence	Young offenders with community sentence and CYF history	Jobseekers in poor health with CYF history	Sole parents not in fulltime employment with CYF history	Long-term disability beneficiaries	In any target population	Not in a target population
Ashburton District	48	63	9	30	27	168	1,629
Auckland (combined)	2,043	2,217	648	1,845	1,266	7,395	94,881
Auckland City	450	468	132	342	318	1,581	28,962
Franklin District	93	114	30	87	66	363	3,561
Manukau City	804	705	159	741	366	2,565	25,923
North Shore City	132	210	66	81	144	597	14,538
Papakura District	159	171	39	183	66	570	3,459
Rodney District	75	135	66	69	72	378	5,286
Waitakere City	330	414	156	342	234	1,341	13,152
Banks Peninsula District	s	6	s	s	6	21	294
Buller District	21	42	9	9	9	81	492
Carterton District	12	15	s	15	s	45	348
Central Hawkes Bay District	18	36	6	18	12	87	600
Central Otago District	24	33	s	6	6	69	690
Christchurch City	651	759	249	381	441	2,259	23,592
Clutha District	30	45	s	18	9	93	861
Dunedin City	261	228	96	111	144	756	10,572
Far North District	204	234	42	150	57	612	2,562
Gisborne District	210	195	27	135	66	579	2,562
Gore District	33	51	s	18	9	111	693
Grey District	33	54	12	27	15	126	771
Hamilton City	324	369	132	300	234	1,242	11,235
Hastings District	231	237	42	171	93	714	3,813
Hauraki District	33	60	18	36	33	168	834
Horowhenua District	75	105	24	99	42	312	1,410
Hurunui District	s	12	s	6	6	33	495
Invercargill City	237	237	27	114	90	645	3,078
Kaikoura District	s	6	s	s	s	12	174
Kaipara District	48	51	6	27	15	135	831

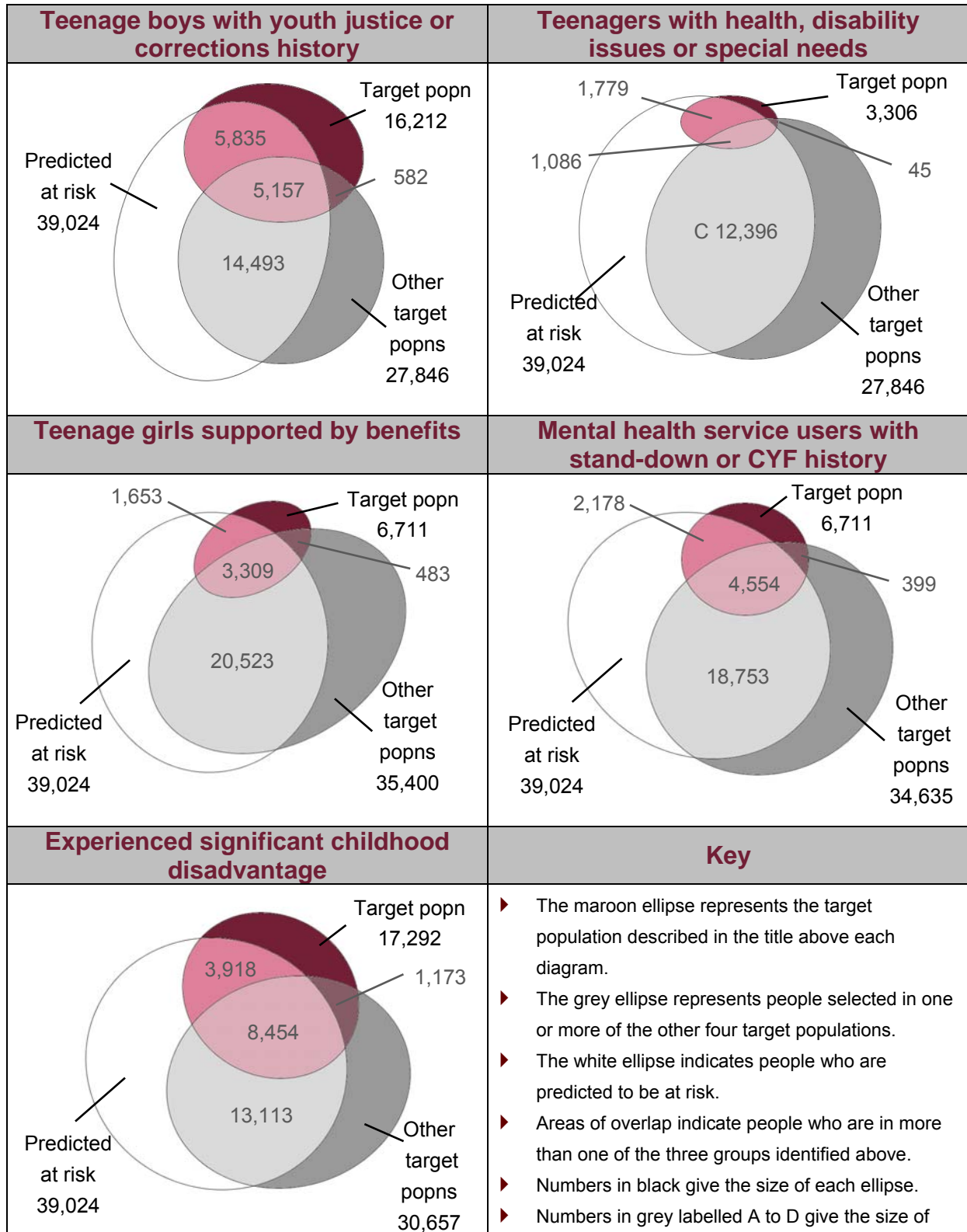
	Young offenders with custodial sentence	Young offenders with community sentence and CYF history	Jobseekers in poor health with CYF history	Sole parents not in fulltime employment with CYF history	Long-term disability beneficiaries	In any target population	Not in a target population
Kapiti Coast District	48	117	15	69	51	270	2,148
Kawerau District	39	39	6	39	15	123	321
Lower Hutt City	156	234	48	171	123	675	6,321
Mackenzie District	s	s	s	s	s	12	153
Manawatu District	39	45	15	36	42	162	1,560
Marlborough District	84	147	24	45	30	306	1,869
Masterton District	42	87	27	51	45	237	1,230
Matamata-Piako District	48	84	18	42	36	210	1,851
Napier City	162	216	51	120	84	567	2,973
Nelson City	138	195	42	72	75	471	2,325
New Plymouth District	183	195	36	114	90	570	3,915
Opotiki District	57	45	s	27	12	129	477
Otorohanga District	18	27	s	15	12	63	531
Palmerston North City	162	207	54	144	132	630	6,342
Porirua City	102	144	21	120	81	423	3,375
Queenstown-Lakes District	24	27	s	s	s	60	1,212
Rangitikei District	36	36	6	21	15	108	783
Rotorua District	231	246	36	171	78	690	3,930
Ruapehu District	45	33	9	33	18	123	654
Selwyn District	24	39	6	6	15	87	2,196
South Taranaki District	78	102	18	51	30	246	1,503
South Waikato District	84	108	15	57	27	270	1,275
South Wairarapa District	12	21	9	12	6	54	366
Southland District	54	60	6	18	9	138	1,497
Stratford District	27	18	s	12	9	66	552
Tararua District	42	39	21	33	24	144	870
Tasman District	54	93	9	39	39	222	2,001
Taupo District	102	111	12	54	27	270	1,629
Tauranga District	288	306	90	186	105	882	5,667
Thames-Coromandel District	30	48	12	30	21	132	984
Timaru District	81	105	24	42	51	279	2,316

	Young offenders with custodial sentence	Young offenders with community sentence and CYF history	Jobseekers in poor health with CYF history	Sole parents not in fulltime employment with CYF history	Long-term disability beneficiaries	In any target population	Not in a target population
Upper Hutt City	66	78	27	48	45	246	2,382
Waikato District	99	105	36	54	48	312	2,658
Waimakariri District	54	90	18	24	39	216	2,472
Waimate District	15	15	s	9	9	42	345
Waipa District	72	66	24	48	63	249	2,586
Wairoa District	33	39	s	27	9	96	378
Waitaki District	54	48	9	21	18	141	903
Waitomo District	24	30	s	18	6	75	498
Wanganui District	171	150	36	102	72	474	2,259
Wellington City	114	141	45	72	144	489	15,786
Western Bay of Plenty District	111	93	21	45	42	282	2,106
Westland District	12	18	s	9	s	39	330
Whakatane District	99	111	15	69	39	309	1,794
Whangarei District	252	327	60	228	93	861	4,092

s = suppressed for confidentiality reasons.

Appendix 6: Target population overlaps – 1990/1991 birth cohort

Appendix 6 Figure 1: Target population overlaps 1990/91 cohort, ages 15 to 19



Appendix 6 Figure 2: Target population overlaps 1990/91 cohort, ages 20 to 24

